(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
27 July 2006 (27.07.2006)

PCT

(10) International Publication Number
WO 2006/078318 A2

(51) International Patent Classification:
A61K 39/02 (2006.01)

(21) International Application Number: PCT/US2005/027239

(22) International Filing Date: 29 July 2005 (29.07.2005)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:

60/592,805	29 July 2004 (29.07.2004)	US
60/609,833	13 September 2004 (13.09.2004)	US
60/616,833	8 October 2004 (08.10.2004)	US
60/633,418	7 December 2004 (07.12.2004)	US
60/640,069	30 December 2004 (30.12.2004)	US
60/660,321	11 March 2005 (11.03.2005)	US
60/673,754	22 April 2005 (22.04.2005)	US
60/693,001	21 June 2005 (21.06.2005)	US
60/695,453	1 July 2005 (01.07.2005)	US
60/697,643	11 July 2005 (11.07.2005)	US

(71) Applicant *(for all designated States except US)*: **CHIRON CORPORATION** [US/US]; 4560 Horton Street, Emeryville, CA 9460-2916 (US).

(72) Inventors; and
(75) Inventors/Applicants *(for US only)*: **TELFORD, John, L.** [IT/IT]; Chiron Vaccines, 1, Via Fiorentina Siena, I-Siena (IT). **GRANDI, Guido** [IT/IT]; CHIRON VACCINES, 1, Via Fiorentia, I-53100 Siena (IT). **LAUER, Peter** [US/US]; 1438 Milvia Street, Berkeley, CA 94709-1917 (US). **MORA, Marirosa** [IT/IT]; CHIRON VACCINES, 1, Via Fiorentina, I-53100 Siena (IT). **ROS, Immaculada, Margarit, Y.** [IT/IT]; CHIRON VACCINES, 1, Via Fiorentina, I-53100 Siena (IT). **MAIONE, Domenico** [IT/IT]; CHIRON VACCINES, 1, Via Fiorentina, I-53100 Siena (IT). **BENSI, Guiliano** [IT/IT]; CHIRON VACCINES, 1, Via Fiorentina, I-53100 Siena (IT). **RINAUDO, Daniela** [IT/IT]; CHIRON VACCINES, 1, Via Fiorentina, I-53100 Siena (IT). **MASIGNANI, Vega** [IT/IT]; CHIRON VACCINES, 1, Via Fiorentina, I-53100 Siena (IT). **BAROCCHI, Michelle** [IT/IT]; CHIRON VACCINES, 1, Via Fiorentina, I-53100 Siena (IT). **RAPPULOI, Rino** [IT/IT]; CHIRON VACINES, 1, Via Fiorentina, I-53100 Siena (IT).

(74) Agent: HALE, Rebecca, M.; CHIRON CORPORATION, 4560 Horton Street, Emeryville, CA 94608-2916 (US).

(81) Designated States *(unless otherwise indicated, for every kind of national protection available)*: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BW, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KM, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NA, NG, NI, NO, NZ, OM, PG, PH, PL, PT, RO, RU, SC, SD, SE, SG, SK, SL, SM, SY, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, YU, ZA, ZM, ZW.

(84) Designated States *(unless otherwise indicated, for every kind of regional protection available)*: ARIPO (BW, GH, GM, KE, LS, MW, MZ, NA, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HU, IE, IS, IT, LT, LU, LV, MC, NL, PL, PT, RO, SE, SI, SK, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:

— *without international search report and to be republished upon receipt of that report*

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

(54) Title: IMMUNOGENIC COMPOSITIONS FOR GRAM POSITIVE BACTERIA SUCH AS STREPTOCOCCUS AGALACTIAE

(57) Abstract: The invention relates to the identification of a new adhesin islands within the genomes of several Group A and Group B Streptococcus serotypes and isolates. The adhesin islands are thought to encode surface proteins which are important in the bacteria's virulence. Thus, the adhesin island proteins of the invention may be used in immunogenic compositions for prophylactic or therapeutic immunization against GAS or GBS infection. For example, the invention may include an immunogenic composition comprising one or more of the discovered adhesin island proteins.

WO 2006/078318 A2

IMMUNOGENIC COMPOSITIONS FOR GRAM POSITIVE BACTERIA SUCH AS *STREPTOCOCCUS AGALACTIAE*

FIELD OF THE INVENTION

The invention relates to the identification of adhesin islands within the genome *Streptococcus agalactiae* ("GBS") and the use of adhesin island amino acid sequences encoded by these adhesin islands in compositions for the treatment or prevention of GBS infection. Similar sequences have been identified in other Gram positive bacteria. The invention further includes immunogenic compositions comprising adhesin island amino acid sequences of Gram positive bacteria for the treatment or prevention of infection of Gram positive bacteria. Preferred immunogenic compositions of the invention include an adhesin island surface protein which may be formulated or purified in an oligomeric or pilus form.

BACKGROUND OF THE INVENTION

GBS has emerged in the last 20 years as the major cause of neonatal sepsis and meningitis that affects 0.5 – 3 per 1000 live births, and an important cause of morbidity among older age groups affecting 5 – 8 per 100,000 of the population. Current disease management strategies rely on intrapartum antibiotics and neonatal monitoring which have reduced neonatal case mortality from >50% in the 1970's to less than 10% in the 1990's. Nevertheless, there is still considerable morbidity and mortality and the management is expensive. 15 – 35% of pregnant women are asymptomatic carriers and at high risk of transmitting the disease to their babies. Risk of neonatal infection is associated with low serotype specific maternal antibodies and high titers are believed to be protective. In addition, invasive GBS disease is increasingly recognized in elderly adults with underlying disease such as diabetes and cancer.

The "B" in "GBS" refers to the Lancefield classification, which is based on the antigenicity of a carbohydrate which is soluble in dilute acid and called the C carbohydrate. Lancefield identified 13 types of C carbohydrate, designated A to O, that could be serologically differentiated. The organisms that most commonly infect humans are found in groups A, B, D, and G. Within group B, strains can be divided into at least 9 serotypes (Ia, Ib, Ia/c, II, III, IV, V, VI, VII and VIII) based on the structure of their polysaccharide capsule. In the past, serotypes Ia, Ib, II, and III were equally prevalent in normal vaginal carriage and early onset sepsis in newborns. Type V GBS has emerged as an important cause of GBS infection in the USA, however, and strains of types VI and VIII have become prevalent among Japanese women.

The genome sequence of a serotype V strain 2603 V/R has been published (See Tettelin *et al.* (2002) *Proc. Natl. Acad. Sci. USA*, 10.1073/pnas.182380799) and various polypeptides for use a vaccine antigens have been identified (WO 02/34771). The vaccines currently in clinical trials, however, are based primarily on polysaccharide antigens. These suffer from serotype-specificity and poor immunogenicity, and so there is a need for effective vaccines against *S.agalactiae* infection.

S. agalactiae is classified as a gram positive bacterium, a collection of about 21 genera of bacteria that colonize humans, have a generally spherical shape, a positive Gram stain reaction and lack endospores. Gram positive bacteria are frequent human pathogens and include *Staphylococcus* (such as *S. aureus*), *Streptococcus* (such as *S. pyogenes* (GBS), *S. pyogenes* (GAS), *S. pneumonaie*, *S. mutans*), *Enterococcus* (such as *E. faecalis* and *E. faecium*), *Clostridium* (such as *C. difficile*), *Listeria* (such as *L. monocytogenes*) and *Corynebacterium* (such as *C. diphtheria*).

It is an object of the invention to provide further and improved compositions for providing immunity against disease and/or infection of Gram positive bacteria. The compositions are based on the identification of adhesin islands within Streptococcal genomes and the use of amino acid sequences encoded by these islands in therapeutic or prophylactic compositions. The invention further includes compositions comprising immunogenic adhesin island proteins within other Gram positive bacteria in therapeutic or prophylactic compositions.

SUMMARY OF THE INVENTION

Applicants have identified a new adhesin island, "GBS Adhesin Island 1", "AI-1" or "GBS AI-1", within the genomes of several Group B Streptococcus serotypes and isolates. This adhesin island is thought to encode surface proteins which are important in the bacteria's virulence. In addition, Applicants have discovered that surface proteins within GBS Adhesin Islands form a previously unseen pilus structure on the surface of GBS bacteria. Amino acid sequences encoded by such GBS Adhesin Islands may be used in immunogenic compositions for the treatment or prevention of GBS infection.

A preferred immunogenic composition of the invention comprises an AI-1 surface protein, such as GBS 80, which may be formulated or purified in an oligomeric (pilus) form. In a preferred embodiment, the oligomeric form is a hyperoligomer. Electron micrographs depicting some of the first visualizations of this pilus structure in a wild type GBS strain are shown in Figures 16, 17, 49, and 50. In addition, Applicants have transformed a GBS strain with a plasmid comprising the AI surface protein GBS 80 which resulted in increased production of that AI surface protein. The electron micrographs of this mutant GBS strain in Figures 13 – 15 reveal long, hyper-oligomeric structures comprising GBS 80 which appear to cover portions of the surface of the bacteria and stretch far out into the supernatant. These hyper-oligomeric pilus structures comprising a GBS AI surface protein may be purified or otherwise formulated for use in immunogenic compositions.

GBS AI-1 comprises a series of approximately five open reading frames encoding for a collection of amino acid sequences comprising surface proteins and sortases ("AI-1 proteins"). Specifically, AI-1 includes polynucleotide sequences encoding for two or more of GBS 80, GBS 104, GBS 52, SAG0647 and SAG0648. One or more of the AI-1 polynucleotide sequences may be replaced by a polynucleotide sequence coding for a fragment of the replaced ORF. Alternatively, one or more of the AI-1 open reading frames may be replaced by a sequence having sequence homology (sequence identity) to the replaced ORF.

AI-1 typically resides on an approximately 16.1 kb transposon-like element frequently inserted into the open reading frame for *trmA*. One or more of the AI-1 surface protein sequences typically include an LPXTG motif (such as LPXTG (SEQ ID NO: 122)) or other sortase substrate motif. The AI surface proteins of the invention may affect the ability of the GBS bacteria to adhere to and invade epithelial cells. AI surface proteins may also affect the ability of GBS to translocate through an epithelial cell layer. Preferably, one or more AI surface proteins are capable of binding to or otherwise associating with an epithelial cell surface. AI surface proteins may also be able to bind to or associate with fibrinogen, fibronectin, or collagen.

The sortase proteins are thought to be involved in the secretion and anchoring of the LPXTG containing surface proteins. AI-1 may encode at least one surface protein. Alternatively, AI-1 may encode at least two surface proteins and at least one sortase. Preferably, AI-1 encodes for at least three surface proteins and at least two sortases. One or more of the surface proteins may include an LPXTG motif or other sortase substrate motif.

The GBS AI-1 protein of the composition may be selected from the group consisting of GBS 80, GBS 104, GBS 52, SAG0647 and SAG0648. GBS AI-1 surface proteins GBS 80 and GBS 104 are preferred for use in the immunogenic compositions of the invention.

In addition to the open reading frames encoding the AI-1 proteins, AI-1 may also include a divergently transcribed transcriptional regulator such as *araC* (*i.e.*, the transcriptional regulator is located near or adjacent to the AI protein open reading frames, but it transcribed in the opposite direction). It is believed that *araC* may regulate the expression of the GBS AI operon. (*See* Korbel et al., Nature Biotechnology (2004) 22(7): 911 – 917 for a discussion of divergently transcribed regulators in *E. coli*).

A second adhesin island, "Adhesin Island-2", "AI-2" or "GBS AI-2", has also been identified in numerous GBS serotypes. Amino acid sequences encoded by the open reading frames of AI-2 may also be used in immunogenic compositions for the treatment or prevention of GBS infection.

GBS AI-2 comprises a series of approximately five open reading frames encoding for a collection of amino acid sequences comprising surface proteins and sortases. Specifically, AI-2 includes open reading frames encoding for two or more of GBS 67, GBS 59, GBS 150, SAG1405, SAG1406, 01520, 01521, 01522, 01523, 01523, 01524 and 01525. The GBS AI-2 sequences may be divided into two subgroups. In one embodiment, AI-2 includes open reading frames encoding for two or more of GBS 67, GBS 59, GBS 150, SAG1405, and SAG1406. This collection of open reading frames may be generally referred to as GBS AI-2 subgroup 1. Alternatively, AI-2 may include open reading frames encoding for two or more of 01520, 01521, 01522, 01523, 01523, 01524 and 01525. This collection of open reading frames may be generally referred to as GBS AI-2 subgroup 2.

One or more of the AI-2 open reading frame polynucleotide sequences may be replaced by a polynucleotide sequence coding for a fragment of the replaced ORF. Alternatively, one or more of the AI-2 open reading frames may be replaced by a sequence having sequence homology (sequence identity) to the replaced ORF.

One or more of the AI-2 surface proteins typically include an LPXTG motif (such as LPXTG (SEQ ID NO: 122)) or other sortase substrate motif. These sortase proteins are thought to be involved in the secretion and anchoring of the LPXTG containing surface proteins. AI-2 may encode for at least one surface protein. Alternatively, AI-2 may encode for at least two surface proteins and at least one sortase. Preferably, AI-2 encodes for at least three surface proteins and at least two sortases. One or more of the surface proteins may include an LPXTG motif.

The AI-2 protein of the composition may be selected from the group consisting of GBS 67, GBS 59, GBS 150, SAG1405, SAG1406, 01520, 01521, 01522, 01523, 01523, 01524 and 01525. AI-2 surface proteins GBS 67, GBS 59, and 01524 are preferred AI-2 proteins for use in the immunogenic compositions of the invention. GBS 67 or GBS 59 is particularly preferred.

GBS AI-2 may also include a divergently transcribed transcriptional regulator such as a RofA like protein (for example *rogB*). As in AI-1, *rogB* is thought to regulate the expression of the AI-2 operon.

The GBS AI proteins of the invention may be used in immunogenic compositions for prophylactic or therapeutic immunization against GBS infection. For example, the invention may include an immunogenic composition comprising one or more GBS AI-1 proteins and one or more GBS AI-2 proteins.

The immunogenic compositions may also be selected to provide protection against an increased range of GBS serotypes and strain isolates. For example, the immunogenic composition may comprise a first and second GBS AI protein, wherein a full length polynucleotide sequence encoding for the first GBS AI protein is not present in a genome comprising a full length polynucleotide sequence encoding for the second GBS AI protein. In addition, each antigen selected for use in the immunogenic compositions will preferably be present in the genomes of multiple GBS serotypes and strain isolates. Preferably, each antigen is presnt in the genomes of at least two (*i.e.*, 3, 4, 5, 6, 7, 8, 9, 10, or more) GBS strain isolates. More preferably, each antigen is present in the genomes of at least two (*i.e.*, at least 3, 4, 5 or more) GBS serotypes.

Within GBS AI-1, Applicants have found that Group B Streptococcus surface exposure of GBS 104 is dependent on the concurrent expression of GBS 80. It is thought that GBS 80 is involved in the transport or localization of GBS 104 to the surface of the bacteria. The two proteins may be oligomerized or otherwise chemically or physically associated. It is possible that this association involves a conformational change in GBS 104 that facilitates its transition to the surface of the GBS bacteria. In addition, one or more AI sortases may also be involved in this surface localization and chemical or physical association. Similar relationships are thought to exist within GBS AI-2. The compositions of the invention may therefore include at least two AI proteins, wherein the two AI proteins are physically or chemically associated. Preferably, the two AI proteins form an oligomer. Preferably, one or more of the AI proteins are in a hyper-oligomeric form. In one embodiment, the associated AI proteins may be purified or isolated from a GBS bacteria or recombinant host cell.

It is also an object of the invention to provide further and improved compositions for providing prophylactic or therapeutic protection against disease and/or infection of Gram positive bacteria. The compositions are based on the identification of adhesin islands within Streptococcal genomes and the use of amino acid sequences encoded by these islands in therapeutic or prophylactic compositions. The invention further includes compositions comprising immunogenic adhesin island proteins within other Gram positive bacteria in therapeutic or prophylactic compositions. Preferred Gram positive adhesin island proteins for use in the invention may be derived from *Staphylococcus* (such as *S. aureus*), *Streptococcus* (such as *S. agalactiae* (GBS), *S. pyogenes* (GAS), *S. pneumonaie*, *S. mutans*), *Enterococcus* (such as *E. faecalis* and *E. faecium*), *Clostridium* (such as *C. difficile*), *Listeria* (such as *L. monocytogenes*) and *Corynebacterium* (such as *C. diphtheria*). Preferably, the Gram positive adhesin island surface proteins are in oligomeric or hyperologimeric form.

For example, Applicants have identified adhesin islands within the genomes of several Group A Streptococcus serotypes and isolates. These adhesion islands are thought to encode surface proteins which are important in the bacteria's virulence, and Applicants have obtained the first electron micrographs revealing the presence of these adhesin island proteins in hyperoligomeric pilus structures on the surface of Group A Streptococcus.

Group A Streptococcus is a human specific pathogen which causes a wide variety of diseases ranging from pharyngitis and impetigo through life threatening invasive disease and necrotizing fasciitis. In addition, post-streptococcal autoimmune responses are still a major cause of cardiac pathology in children.

Group A Streptococcal infection of its human host can generally occur in three phases. The first phase involves attachment and/or invasion of the bacteria into host tissue and multiplication of the bacteria within the extracellular spaces. Generally this attachment phase begins in the throat or the skin. The deeper the tissue level infected, the more severe the damage that can be caused. In the second stage of infection, the bacteria secretes a soluble toxin that diffuses into the surrounding tissue or even systemically through the vasculature. This toxin binds to susceptible host cell receptors and triggers innappropropriate immune responses by these host cells, resulting in pathology. Because the toxin can diffuse throughout the host, the necrosis directly caused by the GAS toxins may be physically located in sites distant from the bacterial infection. The final phase of GAS infection can occur long after the original bacteria have been cleared from the host system. At this stage, the host's previous immune response to the GAS bacteria due to cross reactivity between epitopes of a GAS surface protein, M, and host tissues, such as the heart. A general review of GAS infection can be found in Principles of Bacterial Pathogeneis, Groisman ed., Chapter 15 (2001).

In order to prevent the pathogenic effects associated with the later stages of GAS infection, an effective vaccine against GAS will preferably facilitate host elimination of the bacteria during the initial attachment and invasion stage.

Isolates of Group A Streptococcus are historically classified according to the M surface protein described above. The M protein is surface exposed trypsin-sensitive protein generally comprising two polypeptide chains complexed in an alpha helical formation. The carboxyl terminus is anchored in the cytoplasmic membrane and is highly conserved among all group A streptococci. The amino terminus, which extend through the cell wall to the cell surface, is responsible for the antigenic variability observed among the 80 or more serotypes of M proteins.

A second layer of classification is based on a variable, trypsin-resistant surface antigen, commonly referred to as the T-antigen. Decades of epidemiology based on M and T serological typing have been central to studies on the biological diversity and disease causing potential of Group A Streptococci. While the M-protein component and its inherent variability have been extensively characterized, even after five decades of study, there is still very little known about the structure and variability of T-antigens. Antisera to define T types is commercially available from several sources, including Sevapharma (http://www.sevapharma.cz/en).

The gene coding for one form of T-antigen, T-type 6, from an M6 strain of GAS (D741) has been cloned and characterized and maps to an approximately 11 kb highly variable pathogenicity island. Schneewind et al., J Bacteriol. (1990) 172(6):3310 – 3317. This island is known as the Fibronectin-binding, Collagen-binding T-antigen (FCT) region because it contains, in addition to the T6 coding gene (*tee6*), members of a family of genes coding for Extra Cellular Matrix (ECM) binding proteins. Bessen et al., Infection & Immunity (2002) 70(3):1159-1167. Several of the protein products of this gene family have been shown to directly bind either fibronectin and/or collagen. *See* Hanski et al., Infection & Immunity (1992) 60(12):5119-5125; Talay et al., Infection & Immunity (1992(60(9):3837-3844; Jaffe et al. (1996) 21(2):373-384; Rocha et al., Adv Exp Med Biol. (1997) 418:737-739; Kreikemeyer et al., J Biol Chem (2004) 279(16):15850-15859; Podbielski et al., Mol. Microbiol. (1999) 31(4):1051-64; and Kreikemeyer et al., Int. J. Med Microbiol (2004) 294(2-3):177-88. In some cases direct evidence for a role of these proteins in adhesion and invasion has been obtained.

Applicants raised antiserum against a recombinant product of the *tee6* gene and used it to explore the expression of T6 in M6 strain 2724. In immunoblot of mutanolysin extracts of this strain, the antiserum recognized, in addition to a band corresponding to the predicted molecular mass of the product, very high molecular weight ladders ranging in mobility from about 100 kDa to beyond the resolution of the 3-8% gradient gels used.

This pattern of high molecular weight products is similar to that observed in immunoblots of the protein components of the pili identified in *Streptococcus agalactiae* (described above) and previously in *Corynebacterium diphtheriae*. Electron microscropy of strain M6_2724 with antisera specific for the product of *tee6* revealed abundant surface staining and long pilus like structures extending up to 700 nanometers from the bacterial surface, revealing that the T6 protein, one of the antigens recognized in the original Lancefiled serotyping system, is located within a GAS Adhesin Island (GAS AI-1) and forms long covalently linked pilus structures.

Applicants have identified at least four different Group A Streptococcus Adhesin Islands. While these GAS AI sequences can be identified in numerous M types, Applicants have surprisingly discovered a correlation between the four main pilus subunits from the four different GAS AI types and specific T classifications. While other trypsin-resistant surface exposed proteins are likely also implicated in the T classification designations, the discovery of the role of the GAS adhesin islands (and the associated hyper-oligomeric pilus like structures) in T classification and GAS serotype variance has important implications for prevention and treatment of GAS infections. Applicants have identified protein components within each of the GAS adhesin islands which are associated with the pilus formation. These proteins are believed to be involved in the bacteria's initial adherence mechanisms. Immunological recognition of these proteins may allow the host immune response to slow or prevent the bacteria's transition into the more pathogenic later stages of infection.

In addition, Applicants have discovered that the GBS pili structures appear to be implicated in the formation of biofilms (populations of bacteria growing on a surface, often enclosed in an exopolysaccharide matrix). Biofilms are generally associated with bacterial resistance, as antibiotic treatments and host immune response are frequently unable to erradicate all of the bacteria components of the biofilm. Direction of a host immune response against surface proteins exposed during the first steps of bacterial attachment (i.e., before complete biofilm formation) is preferable.

The invention therefore provides for improved immunogenic compositions against GAS infection which may target GAS bacteria during their initial attachment efforts to the host epithelial cells and may provide protection against a wide range of GAS serotypes. The immunogenic compositions of the invention include GAS AI surface proteins which may be formulated in an oligomeric, or hyperoligomeric (pilus) form. The immunogenic compositions of the invention may include one or more GAS AI surface proteins. The invention also includes combinations of GAS AI surface proteins. Combinations of GAS AI surface proteins may be selected from the same adhesin island or they may be selected from different GAS adhesin islands.

Amino acid sequence encoded by such GAS Adhesin Islands may be used in immunogenic compositions for the treatment or prevention of GAS infection. Preferred immunogenic compositions of the invention comprise a GAS AI surface protein which has been formulated or purified in an oligomeric (pilus) form. In a preferred embodiment, the oligomeric form is a hyperoligomer.

GAS Adhesin Islands generally include a series of open reading frames within a GAS genome that encode for a collection of surface proteins and sortases. A GAS Adhesin Island may encode for an amino acid sequence comprising at least one surface protein. The Adhesin Island, therefore, may encode at least one surface protein. Alternatively, a GAS Adhesin Island may encode for at least two surface proteins and at least one sortase. Preferably, a GAS Adhesin Island encodes for at least three surface proteins and at least two sortases. One or more of the surface proteins may include an LPXTG motif (such as LPXTG (SEQ ID NO: 122)) or other sortase substrate motif. One or more GAS AI surface proteins may participate in the formation of a pilus structure on the surface of the Gram positive bacteria.

GAS Adhesin Islands of the invention preferably include a divergently transcribed transcriptional regulator. The transcriptional regulator may regulate the expression of the GAS AI operon. Examples of transcriptional regulators found in GAS AI sequences include *RofA* and *Nra*.

The GAS AI surface proteins may bind or otherwise adhere to fibrinogen, fibronectin, or collagen. One or more of the GAS AI surface proteins may comprise a fimbrial structural subunit.

One or more of the GAS AI surface proteins may include an LPXTG motif or other sortase substrate motif. The LPXTG motif may be followed by a hydrophobic region and a charged C terminus, which are thought to retard the protein in the cell membrane to facilitate recognition by the membrane-localized sortase. See Barnett, et al., J. Bacteriology (2004) 186 (17): 5865-5875.

GAS AI sequences may be generally categorized as Type 1, Type 2, Type 3, or Type 4, depending on the number and type of sortase sequences within the island and the percentage identity of other proteins (with the exception of RofA and cpa) within the island. Schematics of the GAS adhesin islands are set forth in FIGURE 51A and FIGURE 162. “GAS Adhesin Island-1 or “GAS AI-1” comprises a series of approximately five open reading frames encoding for a collection of amino acid sequences comprising surface proteins and sortases (“GAS AI-1 proteins”). GAS AI-1 preferably comprises surface proteins, a srtB sortase and a rofA divergently transcribed transcriptional regulator. GAS AI-1 surface proteins may include a fibronectin binding protein, a collagen adhesion protein and a fimbrial structural subunit. The fimbrial structural subunit (also known as *tee6*) is thought to form the shaft portion of the pilus like structure, while the collagen adhesion protein (Cpa) is thought to act as an accessory protein facilitating the formation of the pilus structure, exposed on the surface of the bacterial capsule.

Specifically, GAS AI-1 includes polynucleotide sequences encoding for two or more of M6_Spy0157, M6_Spy0158, M6_Spy0159, M6_Spy0160, M6_Spy0161. The GAS AI-1 may also include polynucleotide sequences encoding for any one of CDC SS 410_fimbrial, ISS3650_fimbrial, DSM2071_fimbrial

A preferred immunogenic composition of the invention comprises a GAS AI-1 surface protein which may be formulated or purified in an oligomeric (pilus) form. In a preferred embodiment, the oligomeric form is a hyperoligomer. The immunogenic composition of the invention may alternatively comprise an isolated GAS AI-1 surface protein in oligomeric (pilus) form. The oligomer or hyperoligomeric pilus structures comprising GAS AI-1 surface proteins may be purified or otherwise formulated for use in immunogenic compositions.

One or more of the GAS AI-1 polynucleotide sequences may be replaced by a polynucleotide sequence coding for a fragment of the replaced ORF. Alternatively, one or more of the GAS AI-1 open reading frames may be replaced by a sequence having sequence homology (sequence identity) to the replaced ORF.

One or more of the GAS AI-1 surface proteins typically include an LPXTG motif (such as LPXTG (SEQ ID NO: 122)) or other sortase substrate motif. These sortase proteins are thought to be involved in the secretion and anchoring of the LPXTG containing surface proteins. GAS AI-1 may encode for at least one surface protein. Alternatively, GAS AI-1 may encode for at least two surface proteins and at least one sortase. Preferably, GAS AI-1 encodes for at least three surface proteins and at least two sortases. One or more of the surface proteins may include an LPXTG motif.

GAS AI-1 preferably includes a srtB sortase. GAS srtB sortases may preferably anchor surface proteins with an LPSTG motif (SEQ ID NO: 166), particularly where the motif is followed by a serine.

The GAS AI-1 protein of the composition may be selected from the group consisting of M6_Spy0157, M6_Spy0158, M6_Spy0159, M6_Spy0160 M6_Spy0161, CDC SS 410_fimbrial, ISS3650_fimbrial, and DSM2071_fimbrial. GAS AI-1 surface proteins M6_Spy0157 (a fibronectin binding protein), M6_Spy0159 (a collagen adhesion protein, Cpa), M6_Spy0160 (a fimbrial structural subunit, *tee6*), CDC SS 410_fimbrial (a fimbrial structural subunit), ISS3650_fimbrial (a fimbrial structural subunit), and DSM2071_fimbrial (a fimbrial structural subunit) are preferred GAS AI-1 proteins for use in the immunogenic compositions of the invention. The fimbrial structural subunit tee6 and the collagen adhesion protein Cpa are preferred GAS AI –1 surface proteins. Preferably, each of these GAS AI-1 surface proteins includes an LPXTG sortase substrate motif, such as LPXTG (SEQ ID NO: 122) or LPXSG (SEQ ID NO: 134) (conservative replacement of threonine with serine).

In addition to the open reading frames encoding the GAS AI-1 proteins, GAS AI-1 may also include a divergently transcribed transcriptional regulator such as *rofA* (*i.e.*, the transcriptional regulator is located near or adjacent to the GAS AI protein open reading frames, but it transcribed in the opposite direction).

The GAS AI-1 surface proteins may be used alone, in combination with other GAS AI-1 surface proteins or in combination with other GAS AI surface proteins. Preferably, the immunogenic compositions of the invention include the GAS AI-1 fimbrial structural subunit (*tee6*) and the GAS AI-1 collagen binding protein. Still more preferably, the immunogenic compositions of the invention include the GAS AI-1 fimbrial structural subunit (*tee6*).

A second GAS adhesion island, "GAS Adhesin Island-2" or "GAS AI-2," has also been identified in GAS serotypes. Amino acid sequences encoded by the open reading frames of GAS AI-2 may also be used in immunogenic compositions for the treatment or prevention of GAS infection.

A preferred immunogenic composition of the invention comprises a GAS AI-2 surface protein which may be formulated or purified in an oligomeric (pilus) form. In a preferred embodiment, the oligomeric form is a hyperoligomer. A preferred immunogenic composition of the invention alternatively comprises an isolated GAS AI-2 surface protein in oligomeric (pilus) form. The oligomer or hyperoligomeric pilus structures comprising GAS AI-2 surface proteins may be purified or otherwise formulated for use in immunogenic compositions.

GAS AI-2 comprises a series of approximately eight open reading frames encoding for a collection of amino acid sequences comprising surface proteins and sortases ("GAS AI-2 proteins").

GAS AI-2 preferably comprises surface proteins, a srtB sortase, a srtC1 sortase and a rofA divergently transcribed transcriptional regulator.

Specifically, GAS AI-2 includes polynucleotide sequences encoding for two or more of GAS15, Spy0127, GAS16, GAS17, GAS18, Spy0131, Spy0133, and GAS20.

One or more of the GAS AI-2 polynucleotide sequences may be replaced by a polynucleotide sequence coding for a fragment of the replaced ORF. Alternatively, one or more of the GAS AI-2 open reading frames may be replaced by a sequence having sequence homology (sequence identity) to the replaced ORF.

One or more of the GAS AI-2 surface proteins typically include an LPXTG motif (such as LPXTG (SEQ ID NO: 122)) or other sortase substrate motif. These sortase proteins are thought to be involved in the secretion and anchoring of the LPXTG containing surface proteins. GAS AI-2 may encode for at least one surface protein. Alternatively, GAS AI-2 may encode for at least two surface proteins and at least one sortase. Preferably, GAS AI-2 encodes for at least three surface proteins and at least two sortases. One or more of the surface proteins may include an LPXTG motif.

GAS AI-2 preferably includes a srtB sortase and a srtC1 sortase. As discussed above, GAS srtB sortases may preferably anchor surface proteins with an LPSTG motif (SEQ ID NO: 166), particularly where the motif is followed by a serine. GAS srtC1 sortase may preferentially anchor surface proteins with a V(P/V)PTG (SEQ ID NO:167) motif. GAS srtC1 may be differentially regulated by *rofA*.

The GAS AI-2 protein of the composition may be selected from the group consisting of GAS15, Spy0127, GAS16, GAS17, GAS18, Spy0131, Spy0133, and GAS20. GAS AI-2 surface proteins GAS15 (Cpa), GAS16 (thought to be a fimbrial protein, M1_128), GAS18 (M1_Spy0130), and GAS20 are preferred for use in the immunogenic compositions of the invention. GAS 16 is thought to form the shaft portion of the pilus like structure, while GAS 15 (the collagen adhesion protein Cpa) and GAS 18 are thought to act as accessory proteins facilitating the formation of the pilus structure, exposed on the surface of the bacterial capsule. Preferably, each of these GAS AI-2 surface proteins includes an LPXTG sortase substrate motif, such as LPXTG (SEQ ID NO: 122), VVXTG (SEQ ID NO: 135), or EVXTG (SEQ ID NO: 136).

In addition to the open reading frames encoding the GAS AI-2 proteins, GAS AI-2 may also include a divergently transcribed transcriptional regulator such as *rofA* (*i.e.*, the transcriptional regulator is located near or adjacent to the GAS AI protein open reading frames, but it transcribed in the opposite direction).The GAS AI-2 surface proteins may be used alone, in combination with other GAS AI-2 surface proteins or in combination with other GAS AI surface proteins. Preferably, the immunogenic compositions of the invention include the GAS AI-2 fimbrial protein (GAS 16), the GAS AI-2 collagen binding protein (GAS 15) and GAS 18 (M1_Spy0130). More preferably, the immunogenic compositions of the invention include the GAS AI-2 fimbrial protein (GAS 16).

A third GAS adhesion island, "GAS Adhesin Island-3" or "GAS AI-3," has also been identified in numerous GAS serotypes. Amino acid sequences encoded by the open reading frames of

GAS AI-3 may also be used in immunogenic compositions for the treatment or prevention of GAS infection.

A preferred immunogenic composition of the invention comprises a GAS AI-3 surface protein which may be formulated or purified in an oligomeric (pilus) form. In a preferred embodiment, the oligomeric form is a hyperoligomer. A preferred immunogenic composition of the invention alternatively comprises an isolated GAS AI-3 surface protein in oligomeric (pilus) form. The oligomer or hyperoligomeric pilus structures comprising GAS AI-3 surface proteins may be purified or otherwise formulated for use in immunogenic compositions.GAS AI-3 comprises a series of approximately seven open reading frames encoding for a collection of amino acid sequences comprising surface proteins and sortases ("GAS AI-3 proteins"). GAS AI-3 preferably comprises surface proteins, a srtC2 sortase, and a Negative transcriptional regulator (Nra) divergently transcribed transcriptional regulator. GAS AI-3 surface proteins may include a collagen binding protein, a fimbrial protein, and a F2 like fibronectin-binding protein. GAS AI-3 surface proteins may also include a hypothetical surface protein. The fimbrial protein is thought to form the shaft portion of the pilus like structure, while the collagen adhesion protein (Cpa) and the hypothetical surface protein are thought to act as accessory proteins facilitating the formation of the pilus structure, exposed on the surface of the bacterial capsule. Preferred AI-3 surface proteins include the fimbrial proein, the collagen binding protein and the hypothetical protein. Preferably, each of these GAS AI-3 surface proteins include an LPXTG sortase substrate motif, such as LPXTG (SEQ ID NO: 122), VPXTG (SEQ ID NO: 137), QVXTG (SEQ ID NO: 138) or LPXAG (SEQ ID NO: 139).

Specifically, GAS AI-3 includes polynucleotide sequences encoding for two or more of SpyM3_0098, SpyM3_0099, SpyM3_0100, SpyM3_0101, SpyM3_0102, SpyM3_0103, SpyM3_0104, Sps0100, Sps0101, Sps0102, Sps0103, Sps0104, Sps0105, Sps0106, orf78, orf79, orf80, orf81, orf82, orf83, orf84, spyM18_0126, spyM18_0127, spyM18_0128, spyM18_0129, spyM18_0130, spyM18_0131, spyM18_0132, SpyoM01000156, SpyoM01000155, SpyoM01000154, SpyoM01000153, SpyoM01000152, SpyoM01000151, SpyoM01000150, SpyoM01000149, ISS3040_fimbrial, ISS3776_fimbrial, and ISS4959_fimbrial. In one embodiment, GAS AI-3 may include open reading frames encoding for two or more of SpyM3_0098, SpyM3_0099, SpyM3_0100, SpyM3_0101, SpyM3_0102, SpyM3_0103, and SpyM3_0104. Alternatively, GAS AI-3 may include open reading frames encoding for two or more of Sps0100, Sps0101, Sps0102, Sps0103, Sps0104, Sps0105, and Sps0106. Alternatively, GAS AI-3 may include open reading frames encoding for two or more of orf78, orf79, orf80, orf81, orf82, orf83, and orf84. Alternatively, GAS AI-3 may include open reading frames encoding for two or more of spyM18_0126, spyM18_0127, spyM18_0128, spyM18_0129, spyM18_0130, spyM18_0131, and spyM18_0132. Alternatively, GAS AI-3 may include open reading frames encoding for two or more of SpyoM01000156, SpyoM01000155, SpyoM01000154, SpyoM01000153, SpyoM01000152, SpyoM01000151, SpyoM01000150, and SpyoM01000149. Alternatively, GAS AI-1 may also include polynucleotide sequences encoding for any one of ISS3040_fimbrial, ISS3776_fimbrial, and ISS4959_fimbrial.

One or more of the GAS AI-3 polynucleotide sequences may be replaced by a polynucleotide sequence coding for a fragment of the replaced ORF. Alternatively, one or more of the GAS AI-3 open reading frames may be replaced by a sequence having sequence homology (sequence identity) to the replaced ORF.

One or more of the GAS AI-3 surface proteins typically include an LPXTG motif (such as LPXTG (SEQ ID NO: 122)) or other sortase substrate motif. These sortase proteins are thought to be involved in the secretion and anchoring of the LPXTG containing surface proteins. GAS AI-3 may encode for at least one surface protein. Alternatively, GAS AI-3 may encode for at least two surface proteins and at least one sortase. Preferably, GAS AI-3 encodes for at least three surface proteins and at least two sortases. One or more of the surface proteins may include an LPXTG motif.

GAS AI-3 preferably includes a srtC2 type sortase. GAS srtC2 type sortases may preferably anchor surface proteins with a QVPTG (SEQ ID NO: 140) motif, particularly when the motif is followed by a hydrophobic region and a charged C terminus tail. GAS SrtC2 may be differentially regulated by Nra.

The GAS AI-3 protein of the composition may be selected from the group consisting of SpyM3_0098, SpyM3_0099, SpyM3_0100, SpyM3_0101, SpyM3_0102, SpyM3_0103, SpyM3_0104, Sps0100, Sps0101, Sps0102, Sps0103, Sps0104, Sps0105, Sps0106, orf78, orf79, orf80, orf81, orf82, orf83, orf84, spyM18_0126, spyM18_0127, spyM18_0128, spyM18_0129, spyM18_0130, spyM18_0131, spyM18_0132, SpyoM01000156, SpyoM01000155, SpyoM01000154, SpyoM01000153, SpyoM01000152, SpyoM01000151, SpyoM01000150, SpyoM01000149, ISS3040_fimbrial, ISS3776_fimbrial, and ISS4959_fimbrial. GAS AI-3 surface proteins SpyM3_0098, SpyM3_0100, SpyM3_0102, SpyM3_0104, SPs0100, SPs0102, SPs0104, SPs0106, orf78, orf80, orf82, orf84, spyM18_0126, spyM18_0128, spyM18_0130, spyM18_0132, SpyoM01000155, SpyoM01000153, SpyoM01000151, SpyoM01000149, ISS3040_fimbrial, ISS3776_fimbrial, and ISS4959_fimbrial are preferred GAS AI-3 proteins for use in the immunogenic compositions of the invention.

In addition to the open reading frames encoding the GAS AI-3 proteins, GAS AI-3 may also include a transcriptional regulator such as *Nra.*

GAS AI-3 may also include a LepA putative signal peptidase I protein.

The GAS AI-3 surface proteins may be used alone, in combination with other GAS AI-3 surface proteins or in combination with other GAS AI surface proteins. Preferably, the immunogenic compositions of the invention include the GAS AI-3 fimbrial protein, the GAS AI-3 collagen binding protein, the GAS AI-3 surface protein (such as SpyM3_0102, M3_Sps0104, M5_orf82, or spyM18_0130), and fibronectin binding protein PrtF2. More preferably, the immunogenic compositions of the invention include the GAS AI-3 fimbrial protein, the GAS AI-3 collagen binding protein, and the GAS AI-3 surface protein. Still more preferably, the immunogenic compositions of the invention include the GAS AI-3 fimbrial protein.

Representative examples of the GAS AI-3 fimbrial protein include SpyM3_0100, M3_Sps0102, M5_orf80, spyM18_128, SpyoM01000153, ISS3040_fimbrial, ISS3776_fimbrial, ISS4959_fimbrial.

Representative examples of the GAS AI-3 collagen binding protein include SpyM3_0098, M3_Sps0100, M5_orf 78, spyM18_0126, and SpyoM01000155.

Representative examples of the GAS AI-3 fibronectin binding protein PrtF2 include SpyM3_0104, M3_Sps0106, M5_orf84 and spyM18_0132, and SpyoM01000149.

A fourth GAS adhesion island, "GAS Adhesin Island-4" or "GAS AI-4," has also been identified in GAS serotypes. Amino acid sequences encoded by the open reading frames of GAS AI-4 may also be used in immunogenic compositions for the treatment or prevention of GAS infection.

A preferred immunogenic composition of the invention comprises a GAS AI-4 surface protein which may be formulated or purified in an oligomeric (pilus) form. In a preferred embodiment, the oligomeric form is a hyperoligomer. A preferred immunogenic composition of the invention alternatively comprises an isolated GAS AI-4 surface protein in oligomeric (pilus) form. The oligomer or hyperoligomeric pilus structures comprising GAS AI-3 surface proteins may be purified or otherwise formulated for use in immunogenic compositions. The oligomeric or hyperoligomeric pilus structures comprising GAS AI-4 surface proteins may be purified or otherwise formulated for use in immunogenic compositions.

GAS AI-4 comprises a series of approximately eight open reading frames encoding for a collection of amino acid sequences comprising surface proteins and sortases ("GAS AI-4 proteins"). This GAS adhesin island 4 ("GAS AI-4") comprises surface proteins, a srtC2 sortase, and a RofA regulatory protein. GAS AI-4 surface proteins within may include a fimbrial protein, F1 and F2 like fibronectin-binding proteins, and a capsular polysaccharide adhesion protein (cpa). GAS AI-4 surface proteins may also include a hypothetical surface protein in an open reading frame (orf).

The fimbral protein (EftLSL) is thought to form the shaft portion of the pilus like structure, while the collagen adhesion protein (Cpa) and the hypothetical protein are thought to act as accessory proteins facilitating the formation of the pilus structure, exposed on the surface of the bacterial capsule. Preferably, each of these GAS AI-4 surface proteins include an LPXTG sortase substrate motif, such as LPXTG (SEQ ID NO: 122), VPXTG (SEQ ID NO: 137), QVXTG (SEQ ID NO: 138) or LPXAG (SEQ ID NO: 139).

Specifically, GAS AI-4 includes polynucleotide sequences encoding for two or more of 19224134, 19224135, 19224136, 19224137, 19224138, 19224139, 19224140, and 19224141. A GAS AI-4 polynucleotide may also include polynucleotide sequences encoding for any one of 20010296_fimbrial, 20020069_fimbrial, CDC SS 635_fimbrial, ISS4883_fimbrial, ISS4538_fimbrial. One or more of the GAS AI-4 polynucleotide sequences may be replaced by a polynucleotide sequence coding for a fragment of the replaced ORF. Alternatively, one or more of the GAS AI-4 open reading frames may be replaced by a sequence having sequence homology (sequence identity) to the replaced ORF.

One or more of the GAS AI-4 surface proteins typically include an LPXTG motif (such as LPXTG (SEQ ID NO: 122)) or other sortase substrate motif. These sortase proteins are thought to be involved in the secretion and anchoring of the LPXTG containing surface proteins. GAS AI-4 may encode for at least one surface protein. Alternatively, GAS AI-4 may encode for at least two surface proteins and at least one sortase. Preferably, GAS AI-4 encodes for at least three surface proteins and at least two sortases. One or more of the surface proteins may include an LPXTG motif.

GAS AI-4 includes a SrtC2 type sortase. GAS SrtC2 type sortases may preferably anchor surface proteins with a QVPTG (SEQ ID NO: 140) motif, particularly when the motif is followed by a hydrophobic region and a charged C terminus tail.

The GAS AI-4 protein of the composition may be selected from the group consisting of 19224134, 19224135, 19224136, 19224137, 19224138, 19224139, 19224140, 19224141, 20010296_fimbrial, 20020069_fimbrial, CDC SS 635_fimbrial, ISS4883_fimbrial, and ISS4538_fimbrial. GAS AI-4 surface proteins 19224134, 19224135, 19224137, 19224139, 19224141, 20010296_fimbrial, 20020069_fimbrial, CDC SS 635_fimbrial, ISS4883_fimbrial, ISS4538_fimbrial are preferred proteins for use in the immunogenic compositions of the invention.

In addition to the open reading frames encoding the GAS AI-4 proteins, GAS AI-4 may also include a divergently transcribed transcriptional regulator such as *RofA* (*i.e.*, the transcriptional regulator is located near or adjacent to the AI protein open reading frames, but it transcribed in the opposite direction.

GAS AI-4 may also include a LepA putative signal peptidase I protein and a MsmRL protein. The GAS AI-4 surface proteins may be used alone, in combination with other GAS AI-4 surface proteins or in combination with other GAS AI surface proteins. Preferably, the immunogenic compositions of the invention include the GAS AI-4 fimbrial protein (EftLSL or 20010296_fimbrial, 20020069_fimbrial, CDC SS 635_fimbrial, ISS4883_fimbrial, or ISS4538_fimbrial), the GAS AI-4 collagen binding protein, the GAS AI-4 surface protein (such as M12 isolate A735 orf 2), and fibronectin binding protein PrtF1 and PrtF2. More preferably, the immunogenic compositions of the invention include the GAS AI-4 fimbrial protein, the GAS AI-4 collagen binding protein, and the GAS AI-4 surface protein. Still more preferably, the immunogenic compositions of the invention include the GAS AI-4 fimbrial protein.

The GAS AI proteins of the invention may be used in immunogenic compositions for prophylactic or therapeutic immunization against GAS infection. For example, the invention may include an immunogenic composition comprising one or more GAS AI-1 proteins and one or more of any of GAS AI-2, GAS AI-3, or GAS AI-4 proteins. For example, the invention includes an immunogenic composition comprising at least two GAS AI proteins where each protein is selected from a different GAS adhesin island. The two GAS AI proteins may be selected from one of the following GAS AI combinations: GAS AI-1 and GAS AI-2; GAS AI-1 and GAS AI-3; GAS AI-1 and GAS AI-4; GAS AI-2 and GAS AI-3; GAS AI-2 and GAS AI-4; and GAS AI 3 and GAS AI-4. Preferably the combination includes fimbrial proteins from one or more GAS adhesin islands.

The immunogenic compositions may also be selected to provide protection against an increased range of GAS serotypes and strain isolates. For example, the immunogenic composition may comprise a first and second GAS AI protein, wherein a full length polynucleotide sequence encoding for the first GAS AI protein is not present in a genome comprising a full length polynucleotide sequence encoding for the second GAS AI protein. In addition, each antigen selected for use in the immunogenic compositions will preferably be present in the genomes of multiple GAS serotypes and strain isolates. Preferably, each antigen is present in the genomes of at least two (*i.e.*, 3, 4, 5, 6, 7, 8, 9, 10, or more) GAS strain isolates. More preferably, each antigen is present in the genomes of at least two (*i.e.*, at least 3, 4, 5, or more) GAS serotypes.

Applicants have also identified adhesin islands within the genome of *Streptococcus pneumoniae*. These adhesion islands are thought to encode surface proteins which are important in the bacteria's virulence. Amino acid sequence encoded by such *S. pneumoniae* Adhesin Islands may be used in immunogenic compositions for the treatment or prevention of *S. pneumoniae* infection. Preferred immunogenic compositions of the invention comprise a *S. pneumoniae* AI surface protein which has been formulated or purified in an oligomeric (pilus) form. In a preferred embodiment, the oligomeric form is a hyperoligomer. A preferred immunogenic composition of the invention alternatively comprises an isolated *S. pneumoniae* surface protein in oligomeric (pilus) form. The oligomer or hyperoligomeric pilus structures comprising *S. pneumoniae* surface proteins may be purified or otherwise formulated for use in immunogenic compositions.

The *S. pneumoniae* Adhesin Islands generally include a series of open reading frames within a S. pneumoniae genome that encode for a collection of surface proteins and sortases. A S. pneumoniae Adhesin Island may encode for an amino acid sequence comprising at least one surface protein. Alternatively, the S. pneumoniae Adhesin Island may encode for at least two surface proteins and at least one sortase. Preferably, a S. pneumoniae Adhesin Island encodes for at least three surface proteins and at least two sortases. One or more of the surface proteins may include an LPTXG motif (such as LPXTG (SEQ ID NO: 122)) or other sortase substrate motif. One or more S. pneumoniae AI surface proteins may participate in the formation of a pilus structure on the surface of the S. pneumoniae bacteria.

The S. pneumoniae Adhesin Islands of the invention preferably include a divergently transcribed transcriptional regulator. The transcriptional regulator may regulate the expression of the S. pneumonaie AI operon. An example of a transcriptional regulator found in S. pneumoniae AI sequences is *rlrA.*

A schematic of the organization of a S. pneumoniae AI locus is provided in Figure 137. The locus comprises open reading frames encoding a transcriptional regulator (rlrA), cell wall surface proteins (rrgA, rrgB, rrgC) and sortases (srt B, srtC, srtD).

S. pneumoniae AI sequences may be generally divided into two groups of homology, S. pneuamoniae AI-a and AI-b. S. pneumoniae strains that comprise AI-a include 14 CSR 10, 19A

Hungary 6, 23 F Poland 15, 670, 6B Finland 12, and 6B Spain 2. S. pneumoniae AI strains that comprise AI-b include 19F Taiwan 14, 9V Spain 3, 23F Taiwan 15 and TIGR 4.

S. pneumoniae AI from TIGR4 comprises a series of approximately seven open reading frames encoding for a collection of amino acid sequences comprising surface proteins and sortases ("*S. pneumoniae* AI proteins"). Specifically, *S. pneumoniae* AI from TIGR4 includes polynucleotide sequences encoding for two or more of SP0462, SP0463, SP0464, SP0465, SP0466, SP0467, and SP0468.

One or more of the *S. pneumoniae* AI from TIGR4 polynucleotide sequences may be replaced by a polynucleotide sequence coding for a fragment of the replaced ORF. Alternatively, one or more of the *S. pneumoniae* AI from TIGR4 open reading frames may be replaced by a sequence having sequence homology to the replaced ORF.

S. pneumoniae strain 670 AI comprises a series of approximately seven open reading frames encoding for a collection of amino acid sequences comprising surface proteins and sortases ("*S. pneumoniae* AI proteins"). Specifically, *S. pneumoniae* strain 670 AI includes polynucleotide sequences encoding for two or more of orf1_670, orf3_670, orf4_670, orf5_670, orf6_670, orf7_670, and orf8_670.

One or more of the *S. pneumoniae* strain 670 AI polynucleotide sequences may be replaced by a polynucleotide sequence coding for a fragment of the replaced ORF. Alternatively, one or more of the *S. pneumoniae* strain 670 AI open reading frames may be replaced by a sequence having sequence homology to the replaced ORF.

S. pneumoniae AI from 14 CSR10 comprises a series of approximately seven open reading frames encoding for a collection of amino acid sequences comprising surface proteins and sortases ("*S. pneumoniae* AI proteins"). Specifically, *S. pneumoniae* AI from 14 CSR10 includes polynucleotide sequences encoding for two or more of ORF2_14CSR, ORF3_14CSR, ORF4_14CSR, ORF5_14CSR, ORF6_14CSR, ORF7_14CSR, and ORF8_14CSR.

One or more of the *S. pneumoniae* AI from 14 CSR10 polynucleotide sequences may be replaced by a polynucleotide sequence coding for a fragment of the replaced ORF. Alternatively, one or more of the *S. pneumoniae* AI from 14 CSR10 open reading frames may be replaced by a sequence having sequence homology to the replaced ORF.

S. pneumoniae AI from 19A Hungary 6 comprises a series of approximately seven open reading frames encoding for a collection of amino acid sequences comprising surface proteins and sortases ("*S. pneumoniae* AI proteins"). Specifically, *S. pneumoniae* AI from 19A Hungary 6 includes polynucleotide sequences encoding for two or more of ORF2_19AH, ORF3_19AH, ORF4_19AH, ORF5_19AH, ORF6_19AH, ORF7_19AH, and ORF8_19AH.

One or more of the *S. pneumoniae* AI from 19A Hungary 6 polynucleotide sequences may be replaced by a polynucleotide sequence coding for a fragment of the replaced ORF. Alternatively, one or more of the *S. pneumoniae* AI from 19A Hungary 6 open reading frames may be replaced by a sequence having sequence homology to the replaced ORF.

S. pneumoniae AI from 19F Taiwan 14 comprises a series of approximately seven open reading frames encoding for a collection of amino acid sequences comprising surface proteins and sortases ("*S. pneumoniae* AI proteins"). Specifically, *S. pneumoniae* AI from 19F Taiwan 14 includes polynucleotide sequences encoding for two or more of ORF2_19FTW, ORF3_19FTW, ORF4_19FTW, ORF5_19FTW, ORF6_19FTW, ORF7_19FTW, and ORF8_19FTW.

One or more of the *S. pneumoniae* AI from 19F Taiwan 14 polynucleotide sequences may be replaced by a polynucleotide sequence coding for a fragment of the replaced ORF. Alternatively, one or more of the *S. pneumoniae* AI from 19F Taiwan 14 open reading frames may be replaced by a sequence having sequence homology to the replaced ORF.

S. pneumoniae AI from 23F Poland 16 comprises a series of approximately seven open reading frames encoding for a collection of amino acid sequences comprising surface proteins and sortases ("*S. pneumoniae* AI proteins"). Specifically, *S. pneumoniae* AI from 23F Poland 16 includes polynucleotide sequences encoding for two or more of ORF2_23FP, ORF3_23FP, ORF4_23FP, ORF5_23FP, ORF6_23FP, ORF7_23FP, and ORF8_23FP.

One or more of the *S. pneumoniae* AI from 23F Poland 16 polynucleotide sequences may be replaced by a polynucleotide sequence coding for a fragment of the replaced ORF. Alternatively, one or more of the *S. pneumoniae* AI from 23F Poland 16 open reading frames may be replaced by a sequence having sequence homology to the replaced ORF.

S. pneumoniae AI from 23F Taiwan 15 comprises a series of approximately seven open reading frames encoding for a collection of amino acid sequences comprising surface proteins and sortases ("*S. pneumoniae* AI proteins"). Specifically, *S. pneumoniae* AI from 23F Taiwan 15 includes polynucleotide sequences encoding for two or more of ORF2_23FTW, ORF3_23FTW, ORF4_23FTW, ORF5_23FTW, ORF6_23FTW, ORF7_23FTW, and ORF8_23FTW.

One or more of the *S. pneumoniae* AI from 23F Taiwan 15 polynucleotide sequences may be replaced by a polynucleotide sequence coding for a fragment of the replaced ORF. Alternatively, one or more of the *S. pneumoniae* AI from 23F Taiwan 15 open reading frames may be replaced by a sequence having sequence homology to the replaced ORF.

S. pneumoniae AI from 6B Finland 12 comprises a series of approximately seven open reading frames encoding for a collection of amino acid sequences comprising surface proteins and sortases ("*S. pneumoniae* AI proteins"). Specifically, *S. pneumoniae* AI from 6B Finland 12 includes polynucleotide sequences encoding for two or more of ORF2_6BF, ORF3_6BF, ORF4_6BF, ORF5_6BF, ORF6_6BF, ORF7_6BF, and ORF8_6BF.

One or more of the *S. pneumoniae* AI from 6B Finland 12 polynucleotide sequences may be replaced by a polynucleotide sequence coding for a fragment of the replaced ORF. Alternatively, one or more of the *S. pneumoniae* AI from 6B Finland 12 open reading frames may be replaced by a sequence having sequence homology to the replaced ORF.

S. pneumoniae AI from 6B Spain 2 comprises a series of approximately seven open reading frames encoding for a collection of amino acid sequences comprising surface proteins and sortases

("*S. pneumoniae* AI proteins"). Specifically, *S. pneumoniae* AI from 6B Spain 2 includes polynucleotide sequences encoding for two or more of ORF2_6BSP, ORF3_6BSP, ORF4_6BSP, ORF5_6BSP, ORF6_6BSP, ORF7_6BSP, and ORF8_6BSP.

One or more of the *S. pneumoniae* AI from 6B Spain 2 polynucleotide sequences may be replaced by a polynucleotide sequence coding for a fragment of the replaced ORF. Alternatively, one or more of the *S. pneumoniae* AI from 6B Spain 2 open reading frames may be replaced by a sequence having sequence homology to the replaced ORF.

S. pneumoniae AI from 9V Spain 3 comprises a series of approximately seven open reading frames encoding for a collection of amino acid sequences comprising surface proteins and sortases ("*S. pneumoniae* AI proteins"). Specifically, *S. pneumoniae* AI from 9V Spain 3 includes polynucleotide sequences encoding for two or more of ORF2_9VSP, ORF3_9VSP, ORF4_9VSP, ORF5_9VSP, ORF6_9VSP, ORF7_9VSP, and ORF8_9VSP.

One or more of the *S. pneumoniae* AI from 9V Spain 3 polynucleotide sequences may be replaced by a polynucleotide sequence coding for a fragment of the replaced ORF. Alternatively, one or more of the *S. pneumoniae* AI from 9V Spain 3 open reading frames may be replaced by a sequence having sequence homology to the replaced ORF.

One or more of the *S. pneumoniae* AI surface proteins typically include an LPXTG motif (such as LPXTG (SEQ ID NO: 122)) or other sortase substrate motif. These sortase proteins are thought to be involved in the secretion and anchoring of the LPXTG containing surface proteins. *S. pneumoniae* AI may encode for at least one surface protein. The Adhesin Island, may encode at least one surface protein. Alternatively, *S. pneumoniae* AI may encode for at least two surface proteins and at least one sortase. Preferably, *S. pneumoniae* AI encodes for at least three surface proteins and at least two sortases. One or more of the surface proteins may include an LPXTG motif.

The *S. pneumoniae* AI protein of the composition may be selected from the group consisting of SP0462, SP0463, SP0464, SP0465, SP0466, SP0467, SP0468, orf1_670, orf3_670, orf4_670, orf5_670, orf6_670, orf7_670, orf8_670, ORF2_14CSR, ORF3_14CSR, ORF4_14CSR, ORF5_14CSR, ORF6_14CSR, ORF7_14CSR, ORF8_14CSR, ORF2_19AH, ORF3_19AH, ORF4_19AH, ORF5_19AH, ORF6_19AH, ORF7_19AH, ORF8_19AH, ORF2_19FTW, ORF3_19FTW, ORF4_19FTW, ORF5_19FTW, ORF6_19FTW, ORF7_19FTW, ORF8_19FTW, ORF2_23FP, ORF3_23FP, ORF4_23FP, ORF5_23FP, ORF6_23FP, ORF7_23FP, ORF8_23FP, ORF2_23FTW, ORF3_23FTW, ORF4_23FTW, ORF5_23FTW, ORF6_23FTW, ORF7_23FTW, ORF8_23FTW, ORF2_6BF, ORF3_6BF, ORF4_6BF, ORF5_6BF, ORF6_6BF, ORF7_6BF, ORF8_6BF, ORF2_6BSP, ORF3_6BSP, ORF4_6BSP, ORF5_6BSP, ORF6_6BSP, ORF7_6BSP, ORF8_6BSP, ORF2_9VSP, ORF3_9VSP, ORF4_9VSP, ORF5_9VSP, ORF6_9VSP, ORF7_9VSP and, ORF8_9VSP.

S. pneumoniae AI surface proteins are preferred proteins for use in the immunogenic compositions of the invention. In one embodiment, the compositions of the invention comprise combinations of two or more S pneumoniae AI surface proteins. Preferably such combinations are selected from two or more of the group consisting of SP0462, SP0463, SP0464, orf3_670, orf4_670, orf5_670, ORF3_14CSR, ORF4_14CSR, ORF5_14CSR, ORF3_19AH, ORF4_19AH, ORF5_19AH, ORF3_19FTW, ORF4_19FTW, ORF5_19FTW, ORF3_23FP, ORF4_23FP, ORF5_23FP, ORF3_23FTW, ORF4_23FTW, ORF5_23FTW, ORF3_6BF, ORF4_6BF, ORF5_6BF, ORF3_6BSP, ORF4_6BSP, ORF5_6BSP, ORF3_9VSP, ORF4_9VSP, and ORF5_9VSP.

In addition to the open reading frames encoding the *S. pneumoniae* AI proteins, *S. pneumoniae* AI may also include a transcriptional regulator.

The *S. pneumoniae* AI proteins of the invention may be used in immunogenic compositions for prophylactic or therapeutic immunization against *S. pneumoniae* infection. For example, the invention may include an immunogenic composition comprising one or more *S. pneumoniae* from TIGR4 AI proteins and one or more *S. pneumoniae* strain 670 proteins. The immunogenic composition may comprise one or more AI proteins from any one or more of *S. pneumoniae* strains TIGR4, 19A Hungary 6, 6B Finland 12, 6B Spain 2, 9V Spain 3, 14 CSR 10, 19F Taiwan 14, 23F Taiwan 15, 23F Poland 16, and 670.

The immunogenic compositions may also be selected to provide protection against an increased range of *S. pneumoniae* serotypes and strain isolates. For example, the immunogenic composition may comprise a first and second *S. pneumoniae* AI protein, wherein a full length polynucleotide sequence encoding for the first *S. pneumoniae* AI protein is not present in a genome comprising a full length polynucleotide sequence encoding for the second *S. pneumoniae* AI protein. In addition, each antigen selected for use in the immunogenic compositions will preferably be present in the genomes of multiple *S. pneumoniae* serotypes and strain isolates. Preferably, each antigen is present in the genomes of at least two (*i.e.*, 3, 4, 5, 6, 7, 8, 9, 10, or more) *S. pneumoniae* strain isolates. More preferably, each antigen is present in the genomes of at least two (*i.e.*, at least 3, 4, 5, or more) *S. pneumoniae* serotypes.

The immunogenic compositions may also be selected to provide protection against an increased range of serotypes and strain isolates of a Gram positive bacteria. For example, the immunogenic composition may comprise a first and second Gram positive bacteria AI protein, wherein a full length polynucleotide sequence encoding for the first Gram positive bacteria AI protein is not present in a genome comprising a full length polynucleotide sequence encoding for the second Gram positive bacteria AI protein. In addition, each antigen selected for use in the immunogenic compositions will preferably be present in the genomes of multiple serotypes and strain isolates of the Gram positive bacteria. Preferably, each antigen is present in the genomes of at least two (*i.e.*, 3, 4, 5, 6, 7, 8, 9, 10, or more) Gram positive bacteria strain isolates. More preferably, each antigen is present in the genomes of at least two (*i.e.*, at least 3, 4, 5, or more) Gram positive bacteria serotypes.One or both of the first and second AI proteins may preferably be in oligomeric or hyperoligomeric form.

Adhesin island surface proteins from two or more Gram positive bacterial genus or species may be combined to provide an immunogenic composition for prophylactic or therapeutic treatment of disease or infection of two or more Gram positive bacterial genus or species. Optionally, the adhesin island surface proteins may be associated together in an oligomeric or hyperoligomeric structure.

In one embodiment, the invention comprises adhesin island surface proteins from two or more *Streptococcus* species. For example, the invention includes a composition comprising a GBS AI surface protein and a GAS adhesin island surface protein. As another example, the invention includes a composition comprising a GAS adhesin island surface protein and a *S. pneumoniae* adhesin island surface protein. One or both of the GAS AI surface protein and the *S. pneumoniae* AI surface protein may be in oligomeric or hyperoligomeric form. As a further example, the invention includes a composition comprising a GBS adhesin island surface protein and a *S. pneumoniae* adhesin island surface protein.

In one embodiment, the invention comprises an adhesin island surface protein from two or more Gram positive bacterial genus. For example, the invention includes a composition comprising a *Streptococcus* adhesin island protein and a *Corynebacterium* adhesin island protein. One or more of the Gram positive bacteria AI surface proteins may be in an oligomeric or hyperoligomeric form.

In addition, the AI polynucleotides and amino acid sequences of the invention may also be used in diagnostics to identify the presence or absence of GBS (or a Gram positive bacteria) in a biological sample. They may be used to generate antibodies which can be used to identify the presence of absence of an AI protein in a biological sample or in a prophylactic or therapeutic treatment for GBS (or a Gram positive bacterial) infection. Further, the AI polynucleotides and amino acid sequences of the invention may also be used to identify small molecule compounds which inhibit or decrease the virulence associated activity of the AI.

BRIEF DESCRIPTION OF THE FIGURES

FIGURE 1 presents a schematic depiction of Adhesin Island 1 ("AI-1") comprising open reading frames for GBS 80, GBS 52, SAG0647, SAG0648 and GBS 104.

FIGURE 2 illustrates the identification of AI-1 sequences in several GBS serotypes and strain isolates (GBS serotype V, strain isolate 2603; GBS serotype III, strain isolate nem316; GBS serotype II, strain isolate 18RS21; GBS serotype V, strain isolate CJB111; GBS serotype III, strain isolate COH1 and GBS serotype 1a, strain isolate A909). (An AI-1 was not identified in GBS serotype 1b, strain isolate H36B or GBS serotype 1a, strain isolate 515).

FIGURE 3 presents a schematic depiction of the correlation between AI-1 and the Adhesin Island 2 ("AI-2") within the GBS serotype V, strain isolate 2603 genome. (This AI-2 comprises open reading frames for GBS 67, GBS 59, SAG1406, SAG1405 and GBS 150).

FIGURE 4 illustrates the identification of AI-2 comprising open reading frames encoding for GBS 67, GBS 59, SAG1406, SAG1404 and GBS 150 (or sequences having sequence homology thereto) in several GBS serotypes and strain isolates (GBS serotype V, strain isolate 2603; GBS serotype III, strain isolate NEM316; GBS serotype 1b, strain isolate H36B; GBS serotype V, strain isolate CJB111; GBS serotype II, strain isolate 18RS21; and GBS serotype 1a, strain isolate 515). Figure 4 also illustrates the identification of AI-2 comprising open reading frames encoding for 01520 (a sortase), 01521, 01522 (a sortase), 01523 (spb1), 01524 and 01525 (or sequences having sequence homology thereto).

FIGURE 5 presents data showing that GBS 80 binds to fibronectin and fibrinogen in ELISA.

FIGURE 6 illustrates that all genes in AI-1 are co-transcribed as an operon.

FIGURE 7 presents schematic depictions of in-frame deletion mutations within AI-1.

FIGURE 8 presents FACS data showing that GBS 80 is required for surface localization of GBS 104.

FIGURE 9 presents FACS data showing that sortases SAG0647 and SAG0648 play a semi-redundant role in surface exposure of GBS 80 and GBS 104.

FIGURE 10 presents Western Blots of the in-frame deletion mutants probed with anti-GBS80 and anti-GBS 104 antisera.

FIGURE 11: Electron micrograph of surface exposed pili structures in *Streptococcus agalactiae* containing GBS 80.

FIGURE 12: PHD predicted secondary structure of GBS 067.

FIGURE 13, 14 and 15: Electron micrograph of surface exposed pili structures of strain isolate COH1 of *Streptococcus agalactiae* containing a plasmid insert encoding GBS 80.

FIGURE 16 and 17: Electron micrograph of surface exposed pili structure of wild type strain isolate COH1 of *Streptococcus agalactiae.*

FIGURE 18: Alignment of polynucleotide sequences of AI-1 from serotype V, strain isolates 2603 and CJB111; serotype II, strain isolate 18RS21; serotype III, strain isolates COH1 and NEM316; and serotype 1a, strain isolate A909.

FIGURE 19: Alignment of polynucleotide sequences of AI-2 from serotype V, strain isolates 2603 and CJB111; serotype II, strain isolate 18RS21; serotype 1b, strain isolate H36B; and serotype 1a, strain isolate 515.

FIGURE 20: Alignment of polynucleotide sequences of AI-2 from serotype V, strain isolate 2603 and serotype III, strain isolate NEM316.

FIGURE 21: Alignment of polynucleotide sequences of AI-2 from serotype III, strain isolate COH1 and serotype Ia, strain isolate A909.

FIGURE 22: Alignment of amino acid sequences of AI-1 surface protein GBS 80 from serotype V, strain isolates 2603 and CJB111; serotype 1a, strain isolate A909; serotype III, strain isolates COH1 and NEM316.

FIGURE 23: Alignment of amino acid sequences of AI-1 surface protein GBS 104 from serotype V, strain isolates 2603 and CJB111; serotype III, strain isolates COH1 and NEM316; and serotype II, strain isolate 18RS21.

FIGURE 24: Alignment of amino acid sequences of AI-2 surface protein GBS 067 from serotype V, strain isolates 2603 and CJB111; serotype 1a, strain isolate 515; serotype II, strain isolate 18RS21; serotype Ib, strain isolate H36B; and serotype III, strain isolate NEM316.

FIGURE 25: Illustrates that GBS closely associates with tight junctions and cross the monolayer of ME180 cervical epithelial cells by a paracellular route.

FIGURE 26: Illustrates GBS infection of ME180 cells.

FIGURE 27: Illustrates that GBS 80 recombinant protein does not bind to epithelial cells.

FIGURE 28: Illustrates that deletion of GBS 80 does not effect the capacity of GBS strain 2603 V/R to adhere and invade ME180 cervical epithelial cells.

FIGURE 29: Illustrates binding of recombinant GBS 104 protein to epithelial cells.

FIGURE 30: Illustrates that deletion of GBS 104 in the GBS strain COH1, reduces the capacity of GBS to adhere to ME180 cervical epithelial cells.

FIGURE 31: Illustrates that GBS 80 knockout mutant strain partially loses the ability to translocate through an epithelial cell monolayer.

FIGURE 32: Illustrates that deletion of GBS 104, but not GBS 80, reduces the capacity of GBS to invade J774 macrophage-like cell line.

FIGURE 33: Illustrates that GBS 104 knockout mutant strain translocates through an epithelial monolayer less efficiently than the isogenic wild type.

FIGURE 34: Negative stained electron micrographs of GBS serotype III, strain isolate COH1, containing a plasmid insert to over-express GBS 80.

FIGURE 35: Electron micrographs of surface exposed pili structures on GBS serotype III, strain isolate COH1, containing a plasmid insert to over-express GBS 80, stained with anti-GBS 80 antibodies (visualized with 10 nm gold particles).

FIGURE 36: Electron micrographs of surface exposed pili structures on GBS serotype III, strain isolate COH1, containing a plasmid insert to over-express GBS 80, stained with anti-GBS 80 antibodies (visualized with 10 nm gold particles).

FIGURE 37: Electron micrographs of surface exposed pili structures on GBS serotype III, strain isolate COH1, containing a plasmid insert to over-express GBS 80, stained with anti-GBS 80 antibodies (visualized with 20 n m gold particles).

FIGURE 38: Electron micrographs of surface exposed pili structures on GBS serotype III, strain isolate COH1, containing a plasmid insert to over-express GBS 80, stained with anti-GBS 104 antibodies or preimmune sera (visualized with 10 nm gold particles).

FIGURE 39: Electron micrographs of surface exposed pili structures on GBS serotype III, strain isolate COH1, containing a plasmid insert to over-express GBS 80, stained with anti-GBS 80 antibodies (visualized with 20 nm gold particles) and anti-GBS 104 antibodies (visualized with 10 nm gold particles).

FIGURE 40: Electron micrographs of surface exposed pili structures on GBS serotype III, strain isolate COH1, containing a plasmid insert to over-express GBS 80, stained with anti-GBS 80 antibodies (visualized with 20 nm gold particles) and anti-GBS 104 antibodies (visualized with 10 nm gold particles).

FIGURE 41: Illustrates that GBS 80 is necessary for polymer formation and GBS104 and sortase SAG0648 are necessary for efficient assembly of pili.

FIGURE 42: Illustrates that GBS 67 is part of a second pilus and that GBS 80 is polymerized in strain 515.

FIGURE 43: Illustrates that two macro-molecules are visible in Coh1, one of which is the GBS 80 pilin.

FIGURE 44: Illustrates pilin assembly.

FIGURE 45: Illustrates that GBS 52 is a minor component of the GBS pilus.

FIGURE 46: Illustrates that the pilus is found in the supernatant of a bacterial culture.

FIGURE 47: Illustrates that the pilus is found in the supernatant of bacterial cultures in all phases.

FIGURE 48: Illustrates that in Coh1, only the GBS 80 protein and one sortase (sag0647 or sag0648) is required for polymerization.

FIGURE 49: IEM image of GBS 80 staining of a GBS serotype VIII strain JM9030013 that express pili.

FIGURE 50: IEM image of GBS 104 staining of a GBS serotype VIII strain JM9030013 that express pili.

FIGURE 51A: Schematic depiction of open reading frames comprising a GAS AI-2 serotype M1 isolate, GAS AI-3 serotype M3, M5, M18, and M49 isolates, a GAS AI-4 serotype M12 isolate, and an GAS AI-1 serotype M6 isolate.

FIGURE 51B: Amino acid alignment of SrtC1-type sortase of a GAS AI-2 serotype M1 isolate, SrtC2-type sortases of serotype M3, M5, M18, and M49 isolates, and a SrtC2-type sortase of a GAS AI-4 serotype M12 isolate.

FIGURE 52: Amino acid alignment of the capsular polysaccharide adhesion proteins of GAS AI-4 serotype M12 (A735), GAS AI-3 serotype M5 (Manfredo), *S. pyogenes* strain MGAS315 serotype M3, *S. pyogenes* strain SSI-1 serotype M3, *S. pyogenes* strain MGAS8232 serotype M3, and GAS AI-2 serotype M1.

FIGURE 53: Amino acid alignment of F-like fibronectin-binding proteins of GAS AI-4 serotype M12 (A735) and *S. pyogenes* strain MGAS10394 serotype M6.

FIGURE 54: Amino acid alignment of F2-like fibronectin-binding proteins of GAS AI-4 serotype M12 (A735), *S. pyogenes* strain MGAS8232 serotype M3, GAS AI-3 strain M5 (Manfredo), *S. pyogenes* strain SSI serotype M3, and *S. pyogenes* strain MGAS315 serotype M3.

FIGURE 55: Amino acid alignment of fimbrial proteins of GAS AI-4 serotype M12 (A735), GAS AI-3 serotype M5 (Manfredo), *S. pyogenes* strain MGAS315 serotype M3, *S. pyogenes* strain SSI serotype M3, *S. pyogenes* strain MGAS8232 serotype M3, and *S. pyogenes* M1 GAS serotype M1.

FIGURE 56: Amino acid alignment of hypothetical proteins of GAS AI-4 serotype M12 (A735), *S.* pyogenes strain MGAS315 serotype M3, *S. pyogenes* strain SSI-1 serotype M3, GAS AI-3 serotype M5 (Manfredo), and *S. pyogenes* strain MGAS8232 serotype M3.

FIGURE 57: Results of FASTA homology search for amino acid sequences that align with the collagen adhesion protein of GAS AI-1 serotype M6 (MGAS10394).

FIGURE 58: Results of FASTA homology search for amino acid sequences that align with the fimbrial structural subunit of GAS AI-1 serotype M6 (MGAS10394).

FIGURE 59: Results of FASTA homology search for amino acid sequences that align with the hypothetical protein of GAS AI-2 serotype M1 (SF370).

FIGURE 60: Specifies pilin and E box motifs present in GAS type 3 and 4 adhesin islands.

FIGURE 61: Illustrates that surface expression of GBS 80 protein on GBS strains COH and JM9130013 correlates with formation of pili structures. Surface expression of GBS 80 was determined by FACS analysis using an antibody that cross-hybridizes with GBS 80. Formation of pili structures was determined by immunogold electron microscopy using gold-labelled anti-GBS 80 antibody.

FIGURE 62: Illustrates that surface exposure is capsule-dependent for GBS 322 but not for GBS 80.

FIGURE 63: Illustrates the amino acid sequence identity of GBS 59 proteins in GBS strains.

FIGURE 64: Western blotting of whole GBS cell extracts with anti-GBS 59 antibodies.

FIGURE 65: Western blotting of purified GBS 59 and whole GBS cell extracts with anti-GBS 59 antibodies.

FIGURE 66: FACS analysis of GBS strains CJB111, 7357B, 515 using GBS 59 antiserum.

FIGURE 67: Illustrates that anti-GBS 59 antibodies are opsonic for CJB111 GBS strain serotype V.

FIGURE 68: Western blotting of GBS strain JM9130013 total extracts.

FIGURE 69: Western blotting of GBS stain 515 total extracts shows that GBS 67 and GBS 150 are parts of a pilus.

FIGURE 70: Western blotting of GBS strain 515 knocked out for GBS 67 expression

FIGURE 71: FACS analysis of GBS strain 515 and GBS strain 515 knocked out for GBS 67 expression using GBS 67 and GBS 59 antiserum.

FIGURE 72: Illustrates complementation of GBS 515 knocked out for GBS 67 expression with a construct overexpressing GBS 80.

FIGURE 73: FACS analysis of GAS serotype M6 for spyM6_0159 surface expression.

FIGURE 74: FACS analysis of GAS serotype M6 for spyM6_0160 surface expression.

FIGURE 75: FACS analysis of GAS serotype M1 for GAS 15 surface expression.

FIGURE 76: FACS analysis of GAS serotype M1 for GAS 16 surface expression using a first anti-GAS 16 antiserum.

FIGURE 77: FACS analysis of GAS serotype M1 for GAS 18 surface expression using a first anti-GAS 18 antiserum.

FIGURE 78: FACS analysis of GAS serotype M1 for GAS 18 surface expression using a second anti-GAS 18 antiserum.

FIGURE 79: FACS analysis of GAS serotype M1 for GAS 16 surface expression using a second anti-GAS 16 antisera.

FIGURE 80: FACS analysis of GAS serotype M3 for spyM3_0098 surface expression.

FIGURE 81: FACS analysis of GAS serotype M3 for spyM3_0100 surface expression.

FIGURE 82: FACS analysis of GAS serotype M3 for spyM3_0102 surface expression.

FIGURE 83: FACS analysis of GAS serotype M3 for spyM3_0104 surface expression.

FIGURE 84: FACS analysis of GAS serotype M3 for spyM3_0106 surface expression.

FIGURE 85: FACS analysis of GAS serotype M12 for 19224134 surface expression.

FIGURE 86: FACS analysis of GAS serotype M12 for 19224135 surface expression.

FIGURE 87: FACS analysis of GAS serotype M12 for 19224137 surface expression.

FIGURE 88: FACS analysis of GAS serotype M12 for 19224141 surface expression.

FIGURE 89: Western blot analysis of GAS 15 expression on GAS M1 bacteria.

FIGURE 90: Western blot analysis of GAS 15 expression using GAS 15 immune sera.

FIGURE 91: Western blot analysis of GAS 15 expression using GAS 15 pre-immune sera.

FIGURE 92: Western blot analysis of GAS 16 expression on GAS M1 bacteria.

FIGURE 93: Western blot analysis of GAS 16 expression using GAS 16 immune sera.

FIGURE 94: Western blot analysis of GAS 16 expression using GAS 16 pre-immune sera.

FIGURE 95: Western blot analysis of GAS 18 on GAS M1 bacteria.

FIGURE 96: Western blot analysis of GAS 18 using GAS 18 immune sera.

FIGURE 97: Western blot analysis of GAS 18 using GAS 18 pre-immune sera.

FIGURE 98: Western blot analysis of M6_Spy0159 expression on GAS bacteria.

FIGURE 99: Western blot analysis of 19224135 expression on M12 GAS bacteria.

FIGURE 100: Western blot analysis of 19224137 expression on M12 GAS bacteria.

FIGURE 101: Full length nucleotide sequence of an *S. pneumoniae* strain 670 AI.

FIGURE 102: Western blot analysis of GAS 15, GAS 16, and GAS 18 in GAS M1 strain 2580.

FIGURE 103: Western blot analysis of GAS 15, GAS 16, and GAS 18 in GAS M1 strain 2913.

FIGURE 104: Western blot analysis of GAS 15, GAS 16, and GAS 18 in GAS M1 strain 3280.

FIGURE 105: Western blot analysis of GAS 15, GAS 16, and GAS 18 in GAS M1 strain 3348.

FIGURE 106: Western blot analysis of GAS 15, GAS 16, and GAS 18 in GAS M1 strain 2719.

FIGURE 107: Western blot analysis of GAS 15, GAS 16, and GAS 18 in GAS M1 strain SF370.

FIGURE 108: Western blot analysis of 19224135 and 19224137 in GAS M12 strain 2728.

FIGURE 109: Western blot analysis of 19224139 in GAS M12 strain 2728 using antisera raised against SpyM3_0102.

FIGURE 110: Western blot analysis of M6_Spy0159 and M6_Spy0160 in GAS M6 strain 2724.

FIGURE 111: Western blot analysis of M6_Spy0159 and M6_Spy0160 in GAS M6 strain SF370.

FIGURE 112: Western blot analysis of M6_Spy160 in GAS M6 strain 2724.

FIGURES 113-115: Electron micrographs of surface exposed GAS 15 on GAS M1 strain SF370.

FIGURES 116-121: Electron micrographs of surface exposed GAS 16 on GAS M1 strain SF370.

FIGURES 122-125: Electron micrographs of surface exposed GAS 18 on GAS M1 strain SF370 detected using anti-GAS 18 antisera.

FIGURE 126: IEM image of a hyperoligomer on GAS M1 strain SF370 detected using anti-GAS 18 antisera.

FIGURES 127-132: IEM images of oligomeric and hyperoligomeric structures containing M6_Spy0160 extending from the surface of GAS serotype M6 3650.

FIGURE 133A and B: Western blot analysis of *L. lactis* transformed to express GBS 80 with anti-GBS 80 antiserum.

FIGURES 134: Western blot analyses of *L. lactis* transformed to express GBS AI-1 with anti-GBS 80 antiserum.

FIGURE 135: Ponceau staining of same acrylamide gel as used in Figure 134.

FIGURE 136A: Western blot analysis of sonicated pellets and supernatants of cultured *L. lactis* transformed to express GBS AI-1 polypeptides using anti-GBS 80 antiserum.

FIGURE 136B: Polyacrylamide gel electrophoresis of sonicated pellets and supernatants of cultured *L. lactis* transformed to express GBS AI polypeptides.

FIGURE 137: Depiction of an example *S. pneumoniae* AI locus.

FIGURE 138: Schematic of primer hybridization sites within the *S. pneumoniae* AI locus of FIGURE 137.

FIGURE 139A: The set of amplicons produced from the *S. pneumoniae* strain TIGR4 AI locus.

FIGURE 139B: Base pair lengths of amplicons produced from FIGURE 139A primers in *S. pneumoniae* strain TIGR4.

FIGURE 140: CGH analysis of *S. pneumoniae* strains for the AI locus.

FIGURE 141: Amino acid sequence alignment of polypeptides encoded by AI orf 2 in *S. pneumoniae* AI-positive strains.

FIGURE 142: Amino acid sequence alignment of polypeptides encoded by AI orf 3 in *S. pneumoniae* AI-positive strains.

FIGURE 143: Amino acid sequence alignment of polypeptides encoded by AI orf 4 in *S. pneumoniae* AI-positive strains.

FIGURE 144: Amino acid sequence alignment of polypeptides encoded by AI orf 5 in *S. pneumoniae* AI-positive strains.

FIGURE 145: Amino acid sequence alignment of polypeptides encoded by AI orf 6 in *S. pneumoniae* AI-positive strains.

FIGURE 146: Amino acid sequence alignment of polypeptides encoded by AI orf 7 in *S. pneumoniae* AI-positive strains.

FIGURE 147: Amino acid sequence alignment of polypeptides encoded by AI orf 8 in *S. pneumoniae* AI-positive strains.

FIGURE 148: Diagram comparing amino acid sequences of RrgA in *S. pneumoniae* strains.

FIGURE 149: Amino acid sequence comparison of RrgB *S. pneumoniae* strains.

FIGURE 150A: Sp0462 amino acid sequence.

FIGURE 150B: Primers used to produce a clone encoding the Sp0462 polypeptide.

FIGURE 151A: Schematic depiction of recombinant Sp0462 polypeptide.

FIGURE 151B: Schematic depiction of full-length Sp0462 polypeptide.

FIGURE 152A: Western blot probed with serum obtained from *S. pneumoniae*–infected patients for Sp0462.

FIGURE 152B: Western blot probed with GBS 80 serum for Sp0462.

FIGURE 153A: Sp0463 amino acid sequence.

FIGURE 153B: Primers used to produce a clone encoding the Sp0463 polypeptide.

FIGURE 154A: Schematic depiction of recombinant Sp0463 polypeptide.

FIGURE 154B: Schematic depiction of full-length Sp0463 polypeptide.

FIGURE 155: Western blot detection of recombinant Sp0463 polypeptide.

FIGURE 156: Western blot detection of high molecular weight Sp0463 polymers.

FIGURE 157A: Sp0464 amino acid sequence.

FIGURE 157B: Primers used to produce a clone encoding the Sp0464 polypeptide.

FIGURE 158A: Schematic depiction of recombinant Sp0464 polypeptide.

FIGURE 158B: Schematic depiction of full-length Sp0464 polypeptide.

FIGURE 159: Western blot detection of recombinant Sp0464 polypeptide.

FIGURE 160: Amplification products prepared for production of Sp0462, Sp0463, and Sp0464 clones.

FIGURE 161: Opsonic killing by anti-sera raised against *L. lactis* expressing GBS AI

FIGURE 162: Schematic depicting GAS adhesin islands GAS AI-1, GAS AI-2, GAS AI-3 and GAS AI-4.

FIGURES 163 A-D: Immunoblots of cell-wall fractions of GAS strains with antisera specific for LPXTG proteins of M6_ ISS3650 (A), M1_SF370 (B), M5_ISS4883 (C) and M12_20010296 (D).

FIGURES 163 E-H: Immunoblots of cell-wall fractions of deletion mutants M1_SF370Δ128 (E) M1_SF370Δ130 (F) M1_SF370ΔSrtC1 (G) and the M1_128 deletion strain complemented with plasmid pAM::128 which contains the M1_128 gene (H) with antisera specific for the pilin components of M1_SF370.

FIGURES 163 I-N: Immunogold labelling and transmission electron microscopy of: T6 (I) and Cpa (J) in M6_ISS3650; M1_128 in M1_SF370 (K) and deletion strain M1_SF370Δ128 (N); M5_orf80 in M5_ISS4883 (L); M12_EftLSL.A in M12_20010296 (M). The strains used are indicated below the panels. Bars=200nm.

FIGURE 164: Schematic representation of the FCT region from 7 GAS strains

FIGURES 165 A-H: Flow cytometry of GAS bacteria treated or not with trypsin and stained with sera specific for the major pilus component. Preimmune staining; black lines, untreated bacteria; green lines and trypsin treated bacteria; blue lines. M6_ISS3650 stained with sera which recognize the M6 protein (A) or anti-M6_T6 (B), M1_SF370 stained with anti-M1 (C) or anti-M1_128 (D), M5_ISS4883 stained with anti-PrtF (E) or anti-M5_orf80 (F) and M12_20010296 with anti-M12 (G) or anti-EftLSL.A (H)

FIGURES 166 A-C: Immunoblots of recombinant pilin components with polyvalent Lancefield T-typing sera. The recombinant proteins are shown above the blot and the sera pool used is shown below the blot.

FIGURES 166 D-G: Immunoblots of pilin proteins with monovalent T-typing sera. The recombinant proteins are shown below the blot and the sera used above the blot.

Figure 166 H and I Flow cytometry analysis of strain M1_SF370 (H) and the deletion strain M1_SF370Δ128 (I) with T-typing antisera pool T.

FIGURE 167: Chart describing the number and type of sortase sequences identified within GAS AIs.

FIGURE 168 A: Immunogold-electronmicroscopy of *L. lactis* lacking an expression construct for GBS AI-1 using anti-GBS 80 antibodies.

FIGURE 168 B and C: Immunogold-electronmicroscopy detects GBS 80 in oligomeric (pilus) structures on surface of *L. lactis* transformed to express GBS AI-1

FIGURE 169: FACS analysis detects expression of GBS 80 and GBS 104 on the surface of *L. lactis* transformed to express GBS AI-1.

FIGURE 170: Phase contrast microscopy and immuno-electronmicroscopy shows that expression of GBS AI-1 in *L. lactis* induces *L. lactis* aggregation.

FIGURE 171: Purification of GBS pili from *L. lactis* transformed to express GBS AI-1.

FIGURE 172: Schematic depiction of GAS M6 (AI-1), M1 (AI-2), and M12 (AI-4) adhesin islands and portions of the adhesin islands inserted in the pAM401 construct for expression in *L. lactis*.

FIGURE 173 A-C: Western blot analysis showing assembly of GAS pili in *L. lactis* expressing GAS AI-2 (M1) (A), GAS AI-4 (M12) (B), and GAS AI-1 (M6) (C).

FIGURE 174: FACS analysis of GAS serotype M6 for M6_Spy0157 surface expression.

FIGURE 175: FACS analysis of GAS serotype M12 for 19224139 surface expression.

FIGURE 176 A-E: Immunogold electron microscopy using antibodies against M6_Spy0160 detects pili on the surface of M6 strain 2724.

FIGURE 176 F: Immunogold electron microscopy using antibodies against M6_Spy0159 detects M6_Spy0159 surface expression on M6 strain 2724.

FIGURE 177 A-C: Western blot analysis of M1 strain SF370 GAS bacteria individually deleted for M1_130, SrtC1, or M1_128 using anti-M1_130 serum (A), anti-M1_128 serum (B), and anti-M1_126 serum (C).

FIGURE 178 A-C: Immunogold electron microscopy using antibodies against M1_128 to detect surface expression on wildtype strain SF370 bacteria (A), M1_128 deleted SF370 bacteria (B), and SrtC1 deleted SF370 bacteria (C).

FIGURE 179 A-C: FACS analysis to detect expression of M1_126 (A), M1_128 (B), and M1_130 (C) on the surface of wildtype SF370 GAS bacteria.

FIGURE 179 D-F: FACS analysis to detect expression of M1_126 (D), M1_128 (E), and M1_130 (F) on the surface of M1_128 deleted SF370 GAS bacteria.

FIGURE 179 G-I: FACS analysis to detect expression of M1_126 (G), M1_128 (H), and M1_130 (I) on the surface of SrtC1 deleted SF370 GAS bacteria.

FIGURE 180 A and B: FACS analysis of wildtype (A) and LepA deletion mutant (B) strains of SF370 bacteria for M1 surface expression.

FIGURE 181: Western blot analysis detects high molecular weight polymers in *S. pneumoniae* TIGR4 using anti-RrgB antisera.

FIGURE 182: Detection of high molecular weight polymers in *S. pnuemoniae* rlrA positive strains.

FIGURE 183: Detection of high molecular weight polymers in *S. pneumoniae* TIGR4 by silver staining and Western blot analysis using anti-RrgB antisera.

FIGURE 184: Deletion of *S. pneumoniae* TIGR4 adhesin island sequences interferes with the ability of *S. pneumoniae* to adhere to A549 alveolar cells.

FIGURE 185: Negative staining of *S. pneumoniae* strain TIGR4 showing abundant pili on the bacterial surface.

FIGURE 186: Negative staining of strain TIGR4 deleted for rrgA-srtD adhesin island sequences showing no pili on the bacterial surface

FIGURE 187: Negative staining of the TIGR4 mgrA mutant showing abundant pili on the bacterial surface.

FIGURE 188: Negative staining of the negative control TIGR4 mgrA mutant deleted for adhesin island sequences rrgA-srtD showing no pili on the bacterial surface.

FIGURE 189: Immuno-gold labelling of *S. pneumoniae* strain TIGR4 grown on blood agar solid medium using α-RrgB (5nm) and α-RrgC (10nm). Bar represents 200nm.

FIGURE 190 A and B: Detection of expression and purification of *S. pneumoniae* RrgA protein by SDS-PAGE (A) and Western blot analysis (B).

FIGURE 191: Detection of RrgB by antibodies produced in mice.

FIGURE 192: Detection of RrgC by antibodies produced in mice.

FIGURE 193: Purification of *S. pneumoniae* TIGR 4 pili by a cultivation and digestion method and detection of the purified TIGR4 pili.

FIGURE 194: Purification of *S. pneumoniae* TIGR 4 pili by a sucrose gradient centrifugation method and detection of the purified TIGR4 pili.

FIGURE 195: Purification of *S. pneumoniae* TIGR 4 pili by a gel filtration method and detection of the purified TIGR4 pili.

FIGURE 196: Alignment of full length *S. pneumoniae* adhesin island sequences from ten *S. pneumoniae* strains.

FIGURE 197 A: Schematic of GBS AI-1 coding sequences.

FIGURE 197 B: Nucleotide sequence of intergenic region between AraC and GBS 80 (SEQ ID NO: 273.

FIGURE 197 C: FACS analysis results for GBS 80 expression in GBS strains having different length polyA tracts in the intergenic region between AraC and GBS 80.

FIGURE 198: Table comparing the percent identity of surface proteins encoded by a serotype M6 (harbouring a GAS AI-1) adhesin island relative to other GAS serotypes harbouring an adhesin island.

FIGURE 199: Table comparing the percent identity of surface proteins encoded by a serotype M1 (harbouring a GAS AI-2) adhesin island relative to other GAS serotypes harbouring an adhesin island.

FIGURE 200: Table comparing the percent identity of surface proteins encoded by serotypes M3, M18, M5, and M49 (harbouring GAS AI-3) adhesin islands relative to other GAS serotypes harbouring an adhesin island.

FIGURE 201: Table comparing the percent identity of surface proteins encoded by a serotype M12 (harbouring a GAS AI-1) adhesin island- relative to other GAS serotypes harbouring an adhesin island.

FIGURE 202: GBS 80 recombinant protein does not bind to epithelial cells.

FIGURE 203: Deletion of GBS 80 protein does not affect the ability of GBS to adhere and invade ME180 cervical epithelial cells.

FIGURE 204: GBS 80 binds to extracellular matrix proteins.

FIGURE 205: Deletion of GBS 104 protein, but not GBS 80, reduces the capacity of GBS to invade J774 macrophage-like cells

FIGURE 206: GBS 104 knockout mutant strains of bacteria translocate through an epithelial monolayer less efficiently that the isogenic wild type strain.

FIGURE 207: GBS 80 knockout mutant strains of bacteria partially lose the ability to translocate through an epithelial monolayer.

FIGURE 208: GBS adherence to HUVEC endothelial cells.

FIGURE 209: Strain growth rate of wildtype, GBS 80-deleted, or GBS 104 deleted COH1 GBS.

FIGURE 210: Binding of recombinant GBS 104 protein to epithelial cells by FACS analysis.

FIGURE 211: Deletion of GBS 104 proteinin the GBS strain COH1 reduces the ability of GBS to adhere to ME180 cervical epithelial cells.

FIGURE 212: COH1 strain GBS overexpressing GBS 80 protein has an impaired capacity to translocate through an epithelial monolayer.

FIGURE 213: Scanning electron microscopy shows that overexpression of GBS 80 protein on COH1 strain GBS enhances the capacity of the COH1 bacteria to form microcolonies on epithelial cells.

FIGURE 214: Confocal imaging shows that overexpression of GBS 80 proteins on COH1 strain GBS enhances the capacity of the COH1 bacteria to form microcolonies on epithelial cells.

FIGURE 215: Detection of GBS 59 on the surface of GBS strain 515 by immuno-electron microscopy.

FIGURE 216: Detection of GBS 67 on the surface of GBS strain 515 by immuno-electron microscopy.

FIGURE 217: GBS 67 binds to fibronectin.

FIGURE 218: Western blot analysis shows that deletion of both GBS AI-2 sortase genes abolishes assembly of the pilus.

FIGURE 219: FACS analysis shows that deletion of both GBS AI-2 sortase genes abolishes assembly of the pilus.

FIGURE 220 A-C: Western blot analysis shows that GBS 59, GBS 67, and GBS 150 form high molecular weight complexes.

FIGURE 221 A-C: Western blot analysis shows that GBS 59 is required for polymer formation of GBS 67 and GBS 150.

FIGURE 222: FACS analysis shows that GBS 59 is required for surface exposure of GBS 67.

FIGURE 223: Summary Western blots for detection of GBS 59, GBS 67, or GBS 150 in GBS 515 and GBS 515 mutant strain.

FIGURE 224: Description of GBS 59 allelic variants.

FIGURE 225: GBS 59 is opsonic only against a strain of GBS expressing a homologous GBS 59.

FIGURE 226 A and B: Results of FACS analysis for surface expression of GBS 59 using antibodies specific for different GBS 59isoforms.

FIGURE 227 A and B: Results of FACS analysis for surface expression of GBS 80, GBS 104, GBS 322, GBS 67, and GBS 59 on 41 various strains of GBS bacteria.

FIGURE 228: Results of FACS analysis for surface expression of GBS 80, GBS 104, GBS 322, GBS 67, and GBS 59 on 41 strains of GBS bacteria obtained from the CDC.

FIGURE 229: Expected immunogenicity coverage of different combinations of GBS 80, GBS 104, GBS 322, GBS 67, and GBS 59 across strains of GBS bacteria.

FIGURE 230: GBS 59 opsonophagocytic activity is comparable to that of a mixture of GBS 80, GBS 104, GBS 322 and GBS 67.

FIGURE 231 A-C: Schematic presentation of example hybrid GBS AIs.

FIGURE 232: Schematic presentation of an example hybrid GBS AI.

FIGURE 233 A and B: Western blot and FACS analysis detect expression of GBS 80 and GBS 67 on the surface of L. lactis transformed with a hybrid GBS AI.

FIGURE 234 A-E Hybrid GBS AI cloning strategy.

FIGURE 235: High magnification of *S. pneumoniae* strain TIGR4 pili double labeled with α-RrgB (5nm) and α-RrgC (10nm). Bar represents 100nm.

FIGURE 236: Immuno-gold labeling of the *S. pneumoniae* TIGR4 rrgA-srtD deletion mutant with no visible pili on the surface detectable by α-RrgB- and α-RrgC. Bar represents 200nm.

FIGURE 237: Variability in GBS 67 amino acid sequences between strains 2603 and H36B.

FIGURE 238: Strain variability in GBS 67 amino acid sequences of allele I (2603).

FIGURE 239: Stran variability in GBS 67 amino acid sequence of allele II (H36B).

BRIEF DESCRIPTION OF THE TABLES

TABLE 1: Active Maternal Immunization Assay for fragments of GBS 80

TABLE 2: Passive Maternal Immunization Assay for fragments of GBS 80

TABLE 3: Lethal dose 50% of AI-1 mutants from GBS strain isolate 2603.

TABLE 4: GAS AI-1 sequences from M6 isolate (MGAS10394).

TABLE 5: GAS AI-2 sequences from M1 isolate (SF370).

TABLE 6: GAS AI-3 sequences from M3 isolate (MGAS315).

TABLE 7: GAS AI-3 sequences from M3 isolate (SSI-1).

TABLE 8: GAS AI-3 sequences from M18 isolate (MGAS8232).

TABLE 9: S. pneumoniae AI sequences from TIGR4 sequence.

TABLE 10: GAS AI-3 sequences from M5 isolate (Manfredo).

TABLE 11: GAS AI-4 sequences from M12 isolate (A735).

TABLE 12: Conservation of GBS 80 and GBS 104 amino acid sequences.

TABLE 13: Conservation of GBS 322 and GBS 276 amino acid sequences.

TABLE 14: Active maternal immunization assay for a combination of fragments from GBS 322, GBS 80, GBS 104, and GBS 67.

TABLE 15: Antigen surface exposure of GBS 80, GBS 322, GBS 104, and GBS 67.

TABLE 16: Active maternal immunization assay for each of GBS 80 and GBS 322 antigens.

TABLE 17: Active maternal immunization assay for GBS 59.

TABLE 18: Summary of FACS values for surface expression of spyM6_0159.

TABLE 19: Summary of FACS values for surface expression of spyM6_0160.

TABLE 20: Summary of FACS values for surface expression of GAS 15.

TABLE 21: Summary of FACS values for surface expression of GAS 16.

TABLE 22: Summary of FACS values for surface expression of GAS 16 using a second antisera.

TABLE 23: Summary of FACS values for surface expression of GAS 18.

TABLE 24: Summary of FACS values for surface expression of GAS 18 using a second antisera.

TABLE 25: Summary of FACS values for surface expression of SpyM3_0098.

TABLE 26: Summary of FACS values for surface expression of SpyM3_0100.

TABLE 27: Summary of FACS values for surface expression of SpyM3_0102 in M3 serotypes.

TABLE 28: Summary of FACS values for surface expression of SpyM3_0102 in M6 serotypes.

TABLE 29: Summary of FACS values for surface expression of SpyM3_0104 in M3 serotypes.

TABLE 30: Summary of FACS values for surface expression of SpyM3_0104 in an M12 serotype.

TABLE 31: Summary of FACS values for surface expression of SPs_0106 in M3 serotypes.

TABLE 32: Summary of FACS values for surface expression of SPs_0106 in an M12 serotype.

TABLE 33: Summary of FACS values for surface expression of 19224134 in an M12 serotype.

TABLE 34: Summary of FACS values for surface expression of 19224134 in M6 serotypes.

TABLE 35: Summary of FACS values for surface expression of 19224135 in an M12 serotype.

TABLE 36: Summary of FACS values for surface expression of 19224137 in an M12 serotype.

DETAILED DESCRIPTION OF THE INVENTION

The practice of the present invention will employ, unless otherwise indicated, conventional methods of chemistry, biochemistry, molecular biology, immunology and pharmacology, within the skill of the art. Such techniques are explained fully in the literature. See, e.g., *Remington's Pharmaceutical Sciences*, Mack Publishing Company, Easton, Pa., 19th Edition (1995); *Methods In Enzymology* (S. Colowick and N. Kaplan, eds., Academic Press, Inc.); and *Handbook of Experimental Immunology*, Vols. I-IV (D.M. Weir and C.C. Blackwell, eds., 1986, Blackwell Scientific Publications); Sambrook, et al., *Molecular Cloning: A Laboratory Manual* (2nd Edition, 1989); *Handbook of Surface and Colloidal Chemistry* (Birdi, K.S. ed., CRC Press, 1997); *Short Protocols in Molecular Biology*, 4th ed. (Ausubel et al. eds., 1999, John Wiley & Sons); *Molecular Biology Techniques: An Intensive Laboratory Course*, (Ream et al., eds., 1998, Academic Press); *PCR (Introduction to Biotechniques Series)*, 2nd ed. (Newton & Graham eds., 1997, Springer Verlag);. Peters and Dalrymple, *Fields Virology* (2d ed), Fields et al. (eds.), B.N. Raven Press, New York, NY.

All publications, patents and patent applications cited herein, are hereby incorporated by reference in their entireties.

As used herein, an "Adhesin Island" or "AI" refers to a series of open reading frames within a bacterial genome, such as the genome for Group A or Group B Streptococcus or other gram positive bacteria, that encodes for a collection of surface proteins and sortases. An Adhesin Island may encode for amino acid sequences comprising at least one surface protein. The Adhesin Island may encode at least one surface protein. Alternatively, an Adhesin Island may encode for at least two surface proteins and at least one sortase. Preferably, an Adhesin Island encodes for at least three surface proteins and at least two sortases. One or more of the surface proteins may include an LPXTG motif (such as LPXTG (SEQ ID NO: 122)) or other sortase substrate motif. One or more AI surface proteins may participate in the formation of a pilus structure on the surface of the gram positive bacteria.

Adhesin Islands of the invention preferably include a divergently transcribed transcriptional regulator (*i.e.*, the transcriptional regulator is located near or adjacent to the AI protein open reading frames, but it transcribed in the opposite direction). The transcriptional regulator may regulate the expression of the AI operon.

GBS Adhesin Island 1

As discussed above, Applicants have identified a new adhesin island, "Adhesin Island 1", "AI-1", or "GBS AI-1", within the genomes of several Group B Streptococcus serotypes and isolates. AI-1 comprises a series of approximately five open reading frames encoding for a collection of amino acid sequences comprising surface proteins and sortases ("AI-1 proteins"). Specifically, AI-1 includes open reading frames encoding for two or more (*i.e.*, 2, 3, 4 or 5) of GBS 80, GBS 104, GBS 52, SAG0647 and SAG0648. One or more of the AI-1 open reading frame polynucleotide sequences may be replaced by a polynucleotide sequence coding for a fragment of the replaced ORF. Alternatively, one or more of the AI-1 open reading frames may be replaced by a sequence having sequence homology to the replaced ORF.

A schematic of AI-1 is presented in Figure 1. AI-1 typically resides on an approximately 16.1 kb transposon-like element frequently inserted into the open reading frame for *trmA*. One or more of the AI-1 surface protein sequences typically include an LPXTG motif (such as LPXTG (SEQ ID NO: 122)) motif or other sortase substrate motif. The AI surface proteins of the invention may affect the ability of the GBS bacteria to adhere to and invade epithelial cells. AI surface proteins may also affect the ability of GBS to translocate through an epithelial cell layer. Preferably, one or more AI surface proteins are capable of binding to or otherwise associating with an epithelial cell surface. AI surface proteins may also be able to bind to or associate with fibrinogen, fibronectin, or collagen.

The AI-1 sortase proteins are predicted to be involved in the secretion and anchoring of the LPXTG containing surface proteins. AI-1 may encode for at least one surface protein. Alternatively, AI-1 may encode for at least two surface exposed proteins and at least one sortase. Preferably, AI-1 encodes for at least three surface exposed proteins and at least two sortases. The AI-1 protein preferably includes GBS 80 or a fragment thereof or a sequence having sequence identity thereto.

As used herein, an LPXTG motif represents an amino acid sequence comprising at least five amino acid residues. Preferably, the motif includes a leucine (L) in the first amino acid position, a proline (P) in the second amino acid position, a threonine (T) in the fourth amino acid position and a glycine (G) in the fifth amino acid position. The third position, represented by X, may be occupied by any amino acid residue. Preferably, the X is occupied by lysine (K), Glutamate (E), Asparagine (N), Glutamine (Q) or Alanine (A). Preferably, the X position is occupied by lysine (K). In some embodiments, one of the assigned LPXTG amino acid positions is replaced with another amino acid. Preferably, such replacements comprise conservative amino acid replacements, meaning that the replaced amino acid residue has similar physiological properties to the removed amino acid residue. Genetically encoded amino acids may be divided into four families based on physiological properties: (1) acidic (asparatate and glutamate), (2) basic (lysine, arginine, histitidine), (3) non-polar (alanine, valine, leucine, isoleucine, proline, phenylalanine, methionine, tryptophane) and (4) uncharged polar (glycine, asparagines, glutamine, cysteine, serine, threonine, and tyrosine). Phenylalanine, tryptophan and tyrosine are sometimes classified jointly as aromatic amino acids. For example, it is reasonably predictable that an isolated replacement of a leucine with an isoleucine or valine, an asparate with a glutamate, a threonine with a serine, or a similar conservative replacement of an amino acid with a structurally related amino acid will not have a major effect on the biological activity.

The first amino acid position of the LPXTG motif may be replaced with another amino acid residue. Preferably, the first amino acid residue (leucine) is replaced with an alanine (A), valine (V), isoleucine (I), proline (P), phenylalanine (F), methionine (M), glutamic acid (E), glutamine (Q), or tryptophan (Y) residue. In one preferred embodiment, the first amino acid residue is replaced with an isoleucine (I).

The second amino acid residue of the LPXTG motif may be replaced with another amino acid residue. Preferably, the second amino acid residue praline (P) is replaced with a valine (V) residue.

The fourth amino acid residue of the LPXTG motif may be replaced with another amino acid residue. Preferably, the fourth amino acid residue (threonine) is replaced with a serine (S) or an alanine (A).

In general, an LPXTG motif may be represented by the amino acid sequence XXXXG, in which X at amino acid position 1 is an L, a V, an E, an I, an F, or a Q; X at amino acid position 2 is a P if X at amino acid position 1 is an L, an I, or an F; X at amino acid position 2 is a V if X at amino acid position 1 is a E or a Q; X at amino acid position 2 is a V or a P if X at amino acid position 1 is a V; X at amino acid position 3 is any amino acid residue; X at amino acid position 4 is a T if X at amino acid position 1 is a V, E, I, F, or Q; and X at amino acid position 4 is a T, S, or A if X at amino acid position 1 is an L.

Generally, the LPXTG motif of a GBS AI protein may be represented by the amino acid sequence XPXTG, in which X at amino acid position 1 is L, I, or F, and X at amino acid position 3 is any amino acid residue. Specific examples of LPXTG motifs in GBS AI proteins may include LPXTG (SEQ ID NO: 122) or IPXTG (SEQ ID NO: 133).

As discussed further below, the threonine in the fourth amino acid position of the LPXTG motif may be involved in the formation of a bond between the LPXTG containing protein and a cell wall precursor. Accordingly, in preferred LPXTG motifs, the threonine in the fourth amino acid position is not replaced with another amino acid or, if the threonine is replaced, the replacement amino acid is preferably a conservative amino acid replacement, such as serine.

Instead of an LPXTG motif, the AI surface proteins of the invention may contain alternative sortase substrate motifs such as NPQTN (SEQ ID NO: 142), NPKTN (SEQ ID NO: 168), NPQTG (SEQ ID NO: 169), NPKTG (SEQ ID NO: 170), XPXTGG (SEQ ID NO: 143), LPXTAX (SEQ ID NO: 144), or LAXTGX (SEQ ID NO: 145). (Similar conservative amino acid substitutions can also be made to these membrane motifs).

The AI surface proteins may be covalently attached to the bacterial cell wall by membrane-associated transpeptidases, such as an AI sortase. The sortase may function to cleave the surface protein, preferably between the threonine and glycine residues of an LPXTG motif. The sortase may then assist in the formation of an amide link between the threonine carboxyl group and a cell wall precursor such as lipid II. The precursor can then be incorporated into the peptidoglycan via the transglycoslylation and transpeptidation reactions of bacterial wall synthesis. *See* Comfort et al., Infection & Immunity (2004) 72(5): 2710 – 2722.

The AI surface proteins may be polymerized into pili by sortase-catalysed transpeptidation. (See Figure 44.) Cleavage of AI surface proteins by sortase between the threonine and glycine residues of an LPXTG motif yields a thioester-linked acyl intermediate of sortase. Many AI surface proteins include a pilin motif amino acid sequence which interacts with the sortase and LPXTG amino acid sequence. The first lysine residue in a pilin motif can serve as an amino group acceptor of the cleaved LPXTG motif and thereby provide a covalent linkage between AI subunits to form pili. For example, the pilin motif can make a nucleophilic attack on the acyl enzyme providing a covalent linkage between AI subunits to form pili and regenerate the sortase enzyme. Examples of pilin motifs may include ((YPKN(X_{10})K; SEQ ID NO: 146), (YPKN(X_9)K; SEQ ID NO: 147), (YPK(X_7)K; SEQ ID NO: 148), (YPK(X_{11})K; SEQ ID NO: 149), or (PKN(X_9)K; SEQ ID NO: 150)). Preferably, the AI surface proteins of the invention include a pilin motif amino acid sequence.

Typically, AI surface proteins of the invention will contain an N-terminal leader or secretion signal to facilitate translocation of the surface protein across the bacterial membrane.

Group B Streptococci are known to colonize the urinary tract, the lower gastrointestinal tract and the upper respiratory tract in humans. Electron micrograph images of GBS infection of a cervical epithelial cell line (ME180) are presented in Figure 25. As shown in these images, the bacteria closely associate with tight junctions between the cells and appear to cross the monolayer by a paracellular route. Similar paracellular invasion of ME180 cells is also shown in the contrast images in Figure 26. The AI surface proteins of the invention may effect the ability of the GBS bacteria to adhere to and invade epithelial cells. AI surface proteins may also affect the ability of GBS to translocate through an epithelial cell layer. Preferably, one or more AI surface proteins are capable of binding to or otherwise associating with an epithelial cell surface.

Applicants have discovered that AI-1 surface protein GBS 104 can bind epithelial cells such as ME180 human cervical cells, A549 human lung cells and Caco2 human intestinal cells (See

Figures 29 and 210). Further, deletion of the GBS 104 sequence in a GBS strain reduces the capacity of GBS to adhere to ME180 cervical epithelial cells. (See Figures 30 and 211). Deletion of GBS 104 also reduces the capacity of GBS to invade J774 macrophage-like cells. (See Figures 32 and 205). Deletion of GBS 104 also causes GBS to translocate through epithelial monolayers less efficiently. See Figure 206. GBS 104 protein therefore appears to bind to ME180 epithelial cells and to have a role in adhesion to epithelial cells and macrophage cell lines.

Similar to the GBS bacteria that are deletion mutants for GBS 104, GBS 80 knockout mutant strains also partially lose the ability to translocate through an epithelial monolayer. See Figure 207. Deletion of either GBS 80 or GBS 104 in COH1 cells diminishes adherence to HUVEC endothelial cells. See Figure 208. Deletion of GBS 80 or GBS 104 in COH1 does not, however, affect growth of COH1 either with ME180 cells or in incubation medium (IM). See Figure 209. Both GBS 80 and GBS 104, therefore, appear to be involved in translocation of GBS through epithelial cells.

GBS 80 does not appear to bind to epithelial cells. Incubation of epithelial cells in the presence of GBS 80 protein followed by FACS analysis using an anti-GBS 80 polyclonal antibody did not detect GBS 80 binding to the epithelial cells. See Figure 202. Furthermore, deletion of GBS 80 protein does not affect the ability of GBS to adhere and invade ME180 cervical epithelial cells. See Figure 203

Preferably, one or more of the surface proteins may bind to one or more extracellular matrix (ECM) binding proteins, such as fibrinogen, fibronectin, or collagen. As shown in Figures 5 and 204, and Example 1, GBS 80, one of the AI-1 surface proteins, can bind to the extracellular matrix binding proteins fibronectin and fibrinogen. While GBS 80 protein apparently does not bind to certain epithelial cells or affect the capacity of a GBS bacteria to adhere to or invade cervical epithelial cells (See Figures 27 and 28), removal of GBS 80 from a wild type strain decreases the ability of that strain to translocate through an epithelial cell layer (see Figure 31).

GBS 80 may also be involved in formation of biofilms. COH1 bacteria overexpressing GBS 80 protein have an impaired ability to translocate through an epithelial monolayer. See Figure 212. These COH1 bacteria overexpressing GBS 80 form microcolonies on epithelial cells. See Figures 213 and 214. These microcolonies may be the initiation of biofilm development.

AI Surface proteins may also demonstrate functional homology to previously identified adhesion proteins or extracellular matrix (ECM) binding proteins. For example, GBS 80, a surface protein in AI-1, exhibits some functional homology to *FimA*, a major fimbrial subunit of a Gram positive bacteria *A. naeslundii*. *FimA* is thought to be involved in binding salivary proteins and may be a component in a fimbrae on the surface of *A. naeslundii*. *See* Yeung *et al.* (1997) Infection & Immunity 65:2629-2639; Yeunge *et al* (1998) J. Bacteriol 66:1482-1491; Yeung *et al.* (1988) J. Bacteriol 170:3803 – 3809; and Li *et al.* (2001) Infection & Immunity 69:7224-7233.

A similar functional homology has also been identified between GBS 80 and proteins involved in pili formation in the Gram positive bacteria *Corynebacterium diphtheriae* (SpaA, SpaD, and SpaH). *See*, Ton-That *et al.* (2003) Molecular Microbiology 50(4):1429-1438 and Ton-That *et al.*

(2004) Molecular Microbiology 53(1):251-261. The *C. diphtheriae* proteins all included a pilin motif of WxxxVxVYPK (SEQ ID NO: 151; where x indicates a varying amino acid residue). The lysine (K) residue is particularly conserved in the *C. diphtheriae* pilus proteins and is thought to be involved in sortase catalized oligomerization of the subunits involved in the *C. diphtheriae* pilus structure. (The *C. diphtheriae* pilin subunit SpaA is thought to occur by sortase-catalyzed amide bond cross-linking of adjacent pilin subunits. As the thioester-linked acyl intermediate of sortase requires nucleophilic attack for release, the conserved lysine within the SpaA pilin motif might function as an amino group acceptor of cleaved sorting signals, thereby providing for covalent linkages of the *C. diphtheria* pilin subunits. *See* Figure 6(d) of Ton-That et al., Molecular Microbiology (2003) 50(4):1429-1438.)

In addition, an "E box" comprising a conserved glutamic acid residue has also been identified in the *C. diphtheria* pilin associated proteins as important in *C. diphtheria* pilin assembly. The E box motif generally comprises YxLxETxAPxGY (SEQ ID NO: 152; where x indicates a varying amino acid residue). In particular, the conserved glutamic acid residue within the E box is thought necessary for *C. diphtheria* pilus formation.

Preferably, the AI-1 polypeptides of the immunogenic compositions comprise an E box motif. Some examples of E box motifs in the AI-1 polypeptides may include the amino acid sequences YxLxExxxxxGY (SEQ ID NO: 153), YxLxExxxPxGY (SEQ ID NO: 154), or YxLxETxAPxGY (SEQ ID NO: 152). Specifically, the E box motif of the polypeptides may comprise the amino acid sequences YKLKETKAPEGY (SEQ ID NO: 155), YVLKEIETQSGY (SEQ ID NO: 156), or YKLYEISSPDGY (SEQ ID NO: 157).

As discussed in more detail below, a pilin motif containing a conserved lysine residue and an E box motif containing a conserved glutamic acid residue have both been identified in GBS 80.

While previous publications have speculated that pilus-like structures might be formed on the surface of streptococci, (*see, e.g.*, Ton-That et al., Molecular Microbiology (2003) 50(4): 1429 – 1438), these structures have not been previously visible in negative stain (non-specific) electron micrographs, throwing such speculations into doubt. For example, Figure 34 presents electron micrographs of GBS serotype III, strain isolate COH1 with a plasmid insert to facilitate the overexpression of GBS 80. This EM photo was produced with a standard negative stain – no pilus structures are distinguishable. In addition, the use of such AI surface proteins in immunogenic compositions for the treatment or prevention of infection against a Gram positive bacteria has not been previously described.

Surprisingly, Applicants have now identified the presence of GBS 80 in surface exposed pilus formations visible in electron micrographs. These structures are only visible when the electron micrographs are specifically stained against an AI surface protein such as GBS 80. Examples of these electron micrographs are shown in Figures 11, 16 and 17, which reveal the presence of pilus structures in wild type COH1 *Streptococcus agalactiae*. Other examples of these electron micrographs are shown in Figure 49, which reveals that GBS 80 is associated with pili in a wild type clinical isolate of *S. agalactiae*, JM9030013. (See figure 49.)

Applicants have also constructed mutant GBS strains containing a plasmid comprising the GBS 80 sequence resulting in the overexpression of GBS 80 within this mutant. The electron micrographs of Figures 13 – 15 are also stained against GBS 80 and reveal long, oligomeric structures containing GBS 80 which appear to cover portions of the surface of the bacteria and stretch far out into the supernatant.

In some instances, the formation of pili structures on GBS appears to be correlated to surface expression of GBS 80. Figure 61 provides FAC analysis of GBS 80 surface levels on bacterial strains COH1 and JM9130013 using an anti-GBS 80 antisera. Immunogold electron microscopy of the COH1 and JM9130013 bacteria using anti-GBS 80 antisera demonstrates that JM9130013 bacteria, which have higher values for GBS 80 surface expression, also form longer pili structures.

The surface exposure of GBS 80 on GBS is generally not capsule-dependent. Figure 62 provides FACS analysis of capsulated and uncapsulated GBS analyzed with anti-GBS 80 and anti-GBS 322 antibodies. Surface exposure of GBS 80, unlike GBS 322, is not capsule dependent.

An Adhesin Island surface protein, such as GBS 80 appears to be required for pili formation, as well as an Adhesin Island sortase. Pili are formed in Coh1 bacterial clones that overexpress GBS 80, but lack GBS 104, or one of the AI-1 sortases sag0647 or sag0648. However, pili are not formed in Coh1 bacterial clones that overexpress GBS 80 and lack both sag0647 and sag0648. Thus, for example, it appears that at least GBS 80 and a sortase, sag0647 or sag0648, may be necessary for pili formation. (See Figure 48.) Overexpression of GBS 80 in GBS strain 515, which lacks an AI-1, also assembles GBS 80 into pili. GBS strain 515 contains an AI-2, and thus AI-2 sortases. The AI-2 sortases in GBS strain 515 apparently polymerize GBS 80 into pili. (See Figure 42.) Overexpression of GBS 80 in GBS strain 515 cell knocked out for GBS 67 expression also apparently polymerizes GBS 80 into pili. (See Figure 72.)

While GBS 80 appears to be required for GBS AI-1 pili formation, GBS 104 and sortase SAG0648 appears to be important for efficent AI-1 pili assembly. For example, high-molecular structures are not assembled in isogenic COH1 strains which lack expression of GBS 80 due to gene disruption and are less efficiently assembled in isogenic COH1 strains which lack the expression of GBS 104 (see Figure 41). This GBS strain comprises high molecular weight pili structures composed of covalently linked GBS 80 and GBS 104 subunits. In addition, deleting SAG0648 in COH1 bacteria interferes with assembly of some of the high molecular weight pili structures. Thus, indicating that SAG0648 plays a role in assembly of these pilin species. (See Figure 41).

EM photos confirm the involvement of AI surface protein GBS 104 within the hyperoligomeric structures of a GBS strain adapted for increased GBS 80 expression. (See Figures 34 – 41 and Example 6). In a wild type serotype VIII GBS strain, strain JM9030013, IEM identifies GBS 104 as forming clusters on the bacterial surface. (See Figure 50.)

GBS 52 also appears to be a component of the GBS pili. Immunoblots using an anti-GBS 80 antisera on total cell extracts of Coh1 and a GBS 52 null mutant Coh1 reveal a shift in detected proteins in the Coh1 wild type strain relative to the GBS 52 null mutant Coh1 strain. The shifted proteins were also detected in the wild type Coh1 bacteria with an anti-GBS 52 antisera, indicating that the GBS 52 may be present in the pilus. (See Figure 45.)

In one embodiment, the invention includes a composition comprising oligomeric, pilus-like structures comprising an AI surface protein such as GBS 80. The oligomeric, pilus-like structure may comprise numerous units of AI surface protein. Preferably, the oligomeric, pilus-like structures comprise two or more AI surface proteins. Still more preferably, the oligomeric, pilus-like structure comprises a hyper-oligomeric pilus-like structure comprising at least two (*e.g.*, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 20, 25, 30, 35, 40, 45, 50, 60, 70, 80, 90, 100, 120, 140, 150, 200 or more) oligomeric subunits, wherein each subunit comprises an AI surface protein or a fragment thereof. The oligomeric subunits may be covalently associated via a conserved lysine within a pilin motif. The oligomeric subunits may be covalently associated via an LPXTG motif, preferably, via the threonine amino acid residue.

AI surface proteins or fragments thereof to be incorporated into the oligomeric, pilus-like structures of the invention will preferably include one or both of a pilin motif comprising a conserved lysine residue and an E box motif comprising a conserved glutamic acid residue.

More than one AI surface protein may be present in the oligomeric, pilus-like structures of the invention. For example, GBS 80 and GBS 104 may be incorporated into an oligomeric structure. Alternatively, GBS 80 and GBS 52 may be incorporated into an oligomeric structure, or GBS 80, GBS 104 and GBS 52 may be incorporated into an oligomeric structure.

In another embodiment, the invention includes compositions comprising two or more AI surface proteins. The composition may include surface proteins from the same adhesin island. For example, the composition may include two or more GBS AI-1 surface proteins, such as GBS 80, GBS 104 and GBS 52. The surface proteins may be isolated from Gram positve bacteria or they may be produced recombinantly.

The oligomeric, pilus like structures may be used alone or in the combinations of the invention. In one embodiment, the invention comprises a GBS Adhesin Island protein in oligomeric form, preferably in a hyperoligomeric form. In one embodiment, the invention comprises a composition comprising one or more GBS Adhesin Island 1 ("AI-1") proteins and one or more GBS Adhesin Island 2 ("AI-2") proteins, wherein one or more of the Adhesin Island proteins is in the form of an oligomer, preferably in a hyperoligomeric form.

The oligomeric, pilus-like structures of the invention may be combined with one or more additional GBS proteins. In one embodiment, the oligomeric, pilus-like structures comprise one or more AI surface proteins in combination with a second GBS protein. The second GBS protein may be a known GBS antigen, such as GBS 322 (commonly referred to as "sip") or GBS 276. Nucleotide and amino acid sequences of GBS 322 sequenced from serotype V isolated strain 2603 V/R are set forth in WO 02/35771 as SEQ ID 8539 and SEQ ID 8540 and in the present specification as SEQ ID NOs: 38 and 39. A particularly preferred GBS 322 polypeptide lacks the N-terminal signal peptide, amino acid residues 1-24. An example of a preferred GBS 322 polypeptide is a 407 amino acid fragment and is shown in SEQ ID NO: 40. Examples of preferred GBS 322 polypeptides are further described in PCTUS04/____.____, attorney docket number PP20665.002 filed September 15, 2004, hereby incorporated by reference, published as WO 2005/002619.

Additional GBS proteins which may be combined with the GBS AI surface proteins of the invention are also described in WO 2005/002619. These GBS proteins include GBS 91, GBS 184, GBS 305, GBS 330, GBS 338, GBS 361, GBS 404, GBS 690, and GBS 691.

Additional GBS proteins which may be combined with the GBS AI surface proteins of the invention are described in WO 02/34771.

GBS polysaccharides which may be combined with the GBS AI surface proteins of the invention are described in WO 2004/041157. For example, the GBS AI surface proteins of the invention may be combined with a GBS polysaccharides selected from the group consisting of serotype Ia, Ib, Ia/c, II, III, IV, V, VI, VII and VIII.

The oligomeric, pilus-like structures may be isolated or purified from bacterial cultures in which the bacteria express an AI surface protein. The invention therefore includes a method for manufacturing an oligomeric AI surface antigen comprising culturing a GBS bacterium that expresses the oligomeric AI protein and isolating the expressed oligomeric AI protein from the GBS bacteria. The AI protein may be collected from secretions into the supernatant or it may be purified from the bacterial surface. The method may further comprise purification of the expressed AI protein. Preferably, the AI protein is in a hyperoligomeric form. Macromolecular structures associated with oligomeric pili are observed in the supernatant of cultured GBS strain Coh1. (See Figure 46.) These pili are found in the supernatant at all growth phases of the cultured Coh1 bacteria. (See Figure 47.)

The oligomeric, pilus-like structures may be isolated or purified from bacterial cultures overexpressing an AI surface protein. The invention therefore includes a method for manufacturing an oligomeric Adhesin Island surface antigen comprising culturing a GBS bacterium adapted for increased AI protein expression and isolation of the expressed oligomeric Adhesin Island protein from the GBS bacteria. The AI protein may be collected from secretions into the supernatant or it may be purified from the bacterial surface. The method may further comprise purification of the expressed Adhesin Island protein. Preferably, the Adhesin Island protein is in a hyperoligomeric form.

The GBS bacteria are preferably adapted to increase AI protein expression by at least two (*e.g.*, 2, 3, 4, 5, 8, 10, 15, 20, 25, 30, 35, 40, 45, 50, 60, 70, 80, 90, 100, 125, 150 or 200) times wild type expression levels.

GBS bacteria may be adapted to increase AI protein expression by any means known in the art, including methods of increasing gene dosage and methods of gene upregulation. Such means include, for example, transformation of the GBS bacteria with a plasmid encoding the AI protein. The plasmid may include a strong promoter or it may include multiple copies of the sequence encoding the

AI protein. Optionally, the sequence encoding the AI protein within the GBS bacterial genome may be deleted. Alternatively, or in addition, the promoter regulating the GBS Adhesin Island may be modified to increase expression.

GBS bacteria harbouring a GBS AI-1 may also be adapted to increase AI protein expression by altering the number adenosine nucleotides present at two sites in the intergenic region between AraC and GBS 80. See Figure 197 A, which is a schematic showing the organization of GBS AI-1 and Figure 197 B, which provides the sequence of the intergenic region between AraC and GBS 80 in the AI. The adenosine tracts which applicants have identified as influencing GBS 80 surface expression are at nucleotide positions 187 and 233 of the sequence shown in Figure 197 B (SEQ ID NO: 273). Applicants determined the influence of these adenosine tracts on GBS 80 surface expression in strains of GBS bacteria harboring four adenosines at position 187 and six adenosines at position 233, five adenosines at position 187 and six adenosines position 233, and five adenosines at position 187 and seven adenosines at position 233. FACS analysis of these strains using anti GBS 80 antiserum determined that an intergenic region with five adenosines at position 187 and six adenosines at position 233 had higher expression levels of GBS 80 on their surface than other stains. See Figure 197 C for results obtained from the FACS analysis. Therefore, manipulating the number of adenosines present at positions 187 and 233 of the AraC and GBS 80 intergenic region may further be used to adapt GBS to increase AI protein expression.

The invention further includes GBS bacteria which have been adapted to produce increased levels of AI surface protein. In particular, the invention includes GBS bacteria which have been adapted to produce oligomeric or hyperoligomeric AI surface protein, such as GBS 80. In one embodiment, the Gram positive bacteria of the invention are inactivated or attenuated to permit *in vivo* delivery of the whole bacteria, with the AI surface protein exposed on its surface.

The invention further includes GBS bacteria which have been adapted to have increased levels of expressed AI protein incorporated in pili on their surface. The GBS bacteria may be adapted to have increased exposure of oligomeric or hyperoligomeric AI proteins on its surface by increasing expression levels of a signal peptidase polypeptide. Increased levels of a local signal peptidase expression in Gram positive bacteria (such us LepA in GAS) are expected to result in increased exposure of pili proteins on the surface of Gram positive bacteria. Increased expression of a leader peptidase in GBS may be achieved by any means known in the art, such as increasing gene dosage and methods of gene upregulation. The GBS bacteria adapted to have increased levels of leader peptidase may additionally be adapted to express increased levels of at least one pili protein.

Alternatively, the AI proteins of the invention may be expressed on the surface of a non-pathogenic Gram positive bacteria, such as *Streptococus gordonii* (See, e.g., Byrd et al., "Biological consequences of antigen and cytokine co-expression by recombinant *Streptococcus gordonii* vaccine vectors", Vaccine (2002) 20:2197-2205) or *Lactococcus lactis* (See, e.g., Mannam et al., "Mucosal Vaccine Made from Live, Recombinant Lactococcus lactis Protects Mice against Pharangeal Infection with *Streptococcus pyogenes*" Infection and Immunity (2004) 72(6):3444-3450). As used herein, non-pathogenic Gram positive bacteria refer to Gram positive bacteria which are compatible with a human host subject and are not associated with human pathogenisis. Preferably, the non-pathogenic bacteria are modified to express the AI surface protein in oligomeric, or hyper-oligomeric form. Sequences encoding for an AI surface protein and, optionally, an AI sortase, may be integrated into the non-pathogenic Gram positive bacterial genome or inserted into a plasmid. The non-pathogenic Gram positive bacteria may be inactivated or attenuated to facilitate *in vivo* delivery of the whole bacteria, with the AI surface protein exposed on its surface. Alternatively, the AI surface protein may be isolated or purified from a bacterial culture of the non-pathogenic Gram positive bacteria. For example, the AI surface protein may be isolated from cell extracts or culture supernatants. Alternatively, the AI surface protein may be isolated or purified from the surface of the non-pathogenic Gram positive bacteria.

The non-pathogenic Gram positive bacteria may be used to express any of the Gram positive bacterial Adhesin Island proteins described herein, including proteins from a GBS Adhesin Island, a GAS Adhesin Island, or a S pneumo Adhesin Island. The non-pathogenic Gram positive bacteria are transformed to express an Adhesin Island surface protein. Preferably, the non-pathogenic Gram positive bacteria also express at least one Adhesin Island sortase. The AI transformed non-pathogenic Gram positive bacteria of the invention may be used to prevent or treat infection with a pathogenic Gram positive bacteria, such as GBS, GAS or *Streptococcus pneumoniae*. The non-pathogenic Gram positive bacteria may express the Gram positive bacterial Adheshin Island proteins in oligomeric forms that further comprise adhesin island proteins encoded within the genome of the non-pathogenic Gram positive bacteria.

Applicants modified *L. lactis* to demonstrate that it can express GBS AI polypeptides. *L. lactis* was transformed with a construct encoding GBS 80 under its own promoter and terminator sequences. The transformed *L. lactis* appeared to express GBS 80 as shown by Western blot analysis using anti-GBS 80 antiserum. See lanes 6 and 7 of the Western Blots provided in Figures 133A and 133B (133A and 133B are two different exposures of the same Western blot). See also Example 13.

Applicants also transformed *L. lactis* with a construct encoding GBS AI-1 polypeptides GBS 80, GBS 52, SAG0647, SAG0648, and GBS 104 under the GBS 80 promoter and terminator sequences. These *L. lactis* expressed high molecular weight structures that were immunoreactive with anti-GBS 80 in immunoblots. See Figure 134, lane 2, which shows detection of a GBS 80 monomer and higher molecular weight polymers in total transformed *L. lactis* extracts. Thus, it appeared that *L. lactis* is capable of expressing GBS 80 in oligomeric form. The high molecular weight polymers were not only detected in *L. lactis* extracts, but also in the culture supernatants. See Figure 135 at lane 4. See also Example 14. Thus, the GBS AI polypeptides in oligomeric form can be isolated and purified from either *L. lactis* cell extracts or culture supernatants. These oligomeric forms can, for instance, be isolated from cell extracts or culture supernatants by release by sonication. See Figure 136A and B. See also Figure 171, which shows purification of GBS pili from whole extracts of *L. lactis* expressing the GBS AI-1 following sonication and gel filtration on a Sephacryl HR 400 column.

Furthermore, the *L. lactis* transformed with the construct encoding GBS AI-1 polypeptides GBS 80, GBS 52, SAG0647, SAG0648, and GBS 104 under the GBS 80 promoter and terminator sequences expressed the GBS AI-1 polypeptides on its surface. FACS analysis of these transformed *L. lactis* detected cell surface expression of both GBS 80 and GBS 104. The surface expression levels of GBS 80 and GBS 104 on the transformed *L. lactis* were similar to the surface expression levels of GBS 80 and GBS 104 on GBS strains COH1 and JM9130013, which naturally express GBS AI-1. See Figure 169 for FACS analysis data for *L. lactis* transformed with GBS AI-1 and wildtype JM9130013 bacteria using anti-GBS 80 and GBS 104 antisera. Table 40 provides the results of FACS analysis of transformed *L. lactis*, COH1, and JM9130013 bacteria using anti-GBS 80 and anti-GBS 104 antisera. The numbers provided represent the mean fluorescence value difference calculated for immune versus pre-immune sera obtained for each bacterial strain.

Table 40: FACS analysis of *L. lactis* and GBS bacteria strains expressing GBS AI-1

	Anti-GBS 80 antiserum	Anti-GBS 104 antiserum
GBS AI-1 transformed *L. lactis*	298	251
GBS COH1	305	305
GBS JM9130013	461	355

Immunogold-electronmicroscopy performed with anti-GBS 80 primary antibodies detected the presence of pilus structures on the surface of the *L. lactis* bacteria expressing GBS AI-1, confirming the results of the FACS analysis. See Figure 168 B and C. Interestingly, this expression of GBS pili on the surface of the *L. lactis* induced *L. lactis* aggregation. See Figure 170. Thus, GBS AI polypeptides may also be isolated and purified from the surface of *L. lactis*. The ability of *L. lactis* to express GBS AI polypeptides on its surface also demonstrates that it may be useful as a host to deliver GBS AI antigens.

In fact, immunization of mice with *L. lactis* transformed with GBS AI-1 was protective in a subsequent challenge with GBS. Female mice were immunized with *L. lactis* transformed with GBS AI-1. The immunized female mice were bred and their pups were challenged with a dose of GBS sufficient to kill 90% of non-immunized pups. Detailed protocols for intranasal and subcutaneous immunization of mice with transformed *L. lactis* can be found in Examples 18 and 19, respectively. Table 43 provides data showing that immunization of the female mice with *L. lactis* expressing GBS AI-1 (LL-AI 1) greatly increased survival rate of challenged pups relative to both a negative PBS control (PBS) and a negative *L. lactis* control (LL 10 E9, which is wild type *L. lactis* not transformed to express GBS AI-1).

Table 43: Protection of Mice Immunized with *L. lactis* expressing GBS AI-1

Immunization Route	Antigen	Alive/Treated	Survival %	Survival % Range	p value
Intraperitoneum	Recombinant GBS 80	16/18	89	80-100	<0.001
Subcutaneous	LL-AI 1 10 E9	40/49	82	70-90	<0.001
	LL-AI 1 10 E10	50/60	83	60-100	<0.001
	PBS	4/30	13	0-30	
	LL 10 E9	3/57	5	0-20	
Intranasal	LL-AI 1 10 E9	22/60	37	0-100	0.02

	LL-AI 1 10 E10	31/49	63	30-90	<0.001
	LL 10 E9	2/27	7	0-20	

Table 51 provides further evidence that immunization of mice with *L. lactis* transformed with GBS AI-1 is protective against GBS.

Table 51: Further Protection of Mice Immunized with *L. lactis* expressing GBS AI-1

Antigen	Immunization route	Alive/Treated	Survival % (Pval <0.0000001)
Recombinant GBS 80	IP	48/50	**92**
Recombinant GBS 80	SC	21/30	**70**
L.lactis+AI1 10^6 cfu	SC	6/66	9
L.lactis+AI1 10^7 cfu	SC	47/70	**73**
L.lactis+AI1 10^8 cfu	SC	116/153	**76**
L.lactis+AI1 10^9 cfu	SC	98/118	**83**
L.lactis+AI1 10^{10} cfu	SC	107/129	**83**
L.lactis 10^{10} cfu	SC	4/83	5
PBS	SC	6/110	5
L.lactis+AI1 10^{10} cfu	IN	51/97	**52**
L.lactis 10^{11} cfu	IN	1/40	7
PBS	IN	0/37	0

Protection of immunized mice with *L. lactis* expressing the GBS AI-1 is at least partly due to a newly raised antibody response. Table 46 provides anti-GBS 80 antibody titers detected in serum of the mice immunized with *L. lactis* expressing the GBS AI-1 as described above. Mice immunized with *L. lactis* expressing the GBS AI-1 have anti-GBS 80 antibody titres, which are not observed in mice immunized with *L. lactis* not transformed to express the GBS AI-1. Further, as expected from the survival data, mice subcutaneously immunized with *L. lactis* transformed to express the GBS AI-1 have significantly higher serum anti-GBS 80 antibody titers than mice intranasally immunized with *L. lactis* transformed to express the GBS AI-1.

Table 46: Antibody Responses against GBS 80 in Serum of Mice Immunized with *L. lactis* Expressing GBS AI-1

Antigen	Ab Titre Obtained Following		
	Subcutaneous Immunization	Intranasal Immunization	Intraperitoneal Immunization
LL 10 E9	0	0	
LL-AI 1 10 E9	14000	50	
LL-AI 1 10 E10	25000	406	
Recombinant GBS 80			120000

Anti-GBS 80 antibodies of the IgA isotype were specifically detected in various body fluids of the mice subcutaneously or intranasally immunized with *L. lactis* expressing the GBS AI-1.

Table 47: Anti-GBS 80 IgA Antibodies Detected in Mouse Tissues Following Immunization with *L. lactis* Expressing GBS AI-1

Antigen	Immunization route	Anti-GBS 80 IgA Antibodies Detected in		
		Serum	Vaginal Wash	Nasal Wash
LL 10 E9		0	0	0
LL-AI 1	Subcutaneous	0	25	20
LL-AI 1	Intranasal	140	0	150
GBS 80	Intraperitoneal	60	0	

Furthermore, opsonophagocytosis assays also demonstrated that at least some of the antiserum produced against the *L. lactis* expressing GBS AI 1 is opsonic for GBS. See Figure 161.

To obtain protection of against GBS across a greater number of strains and serotypes, it is possible to transform *L. lactis* with a recombinant GBS AI encoding both GBS AI-1 and AI-2, *i.e.*, a hybrid GBS AI. By way of example, a hybrid GBS AI may be a GBS AI-1 with a replacement of the GBS 104 gene with a GBS 67 gene. A schematic of such a hybrid GBS AI is depicted in Figure 231 A. A hybrid GBS AI may alternatively be a GBS AI-1 with a replacement of the GBS 52 gene with a GBS 59 gene. See the schematic at Figure 231 B. Alternatively, a hybrid GBS AI may be a GBS AI-1 with a substitution of a GBS 59 polypeptide for the GBS 52 gene and a substitution of the GBS 104 gene for genes encoding GBS 59 and the two GBS AI-2 sortases. Another example of a hybrid GBS AI is a GBS AI-1 with the substitution of a GBS 59 gene for the GBS 52 gene and a GBS 67 for the GBS 104 gene. See the schematic at Figure 232. A further example of a hybrid GBS AI is a GBS AI-1 having a GBS 59 gene and genes encoding the GBS AI-2 sortases in place of the GBS 52 gene. Yet another example of a hybrid GBS AI is a GBS AI-1 with a substitution of either GBS 52 or GBS 104 with a fusion protein comprising GBS 322 and one of GBS 59, GBS 67, or GBS 150. Some of these hybrid GBS AIs may be prepared as briefly outlined in Figure 234 A-F.

Applicants have prepared a hybrid GBS AI having a GBS AI-1 sequence with a substitution of a GBS 67 coding sequence for the GBS 104 gene as depicted in Figure 231 A. Transformation of *L. lactis* with the hybrid GBS AI-1 resulted in *L. lactis* expression of high molecular weight polymers containing the GBS 80 and GBS 67 proteins. See Figure 233 A, which provides Western blot analysis of *L. lactis* transformed with the hybrid GBS AI depicted in Figure 231 A. When *L. lactis* transformed with the hybrid GBS AI were probed with antibodies to GBS 80 or GBS 67, high molecular weight structures were detected. See lanes labelled LL + a) in both the α-80 and α-67 immunoblots. The GBS 80 and GBS 67 proteins were confirmed to be present on the surface of *L. lactis* by FACS analysis. See Figure 233 B, which shows a shift in fluorescence when GBS 80 and GBS 67 antibodies are used to detect GBS 80 and GBS 67 surface expression. The same shifts in fluorescence were not observed in *L. lactis* control cells, cells not transformed with the hybrid GBS AI.

Alternatively, the oligomeric, pilus-like structures may be produced recombinantly. If produced in a recombinant host cell system, the AI surface protein will preferably be expressed in coordination with the expression of one or more of the AI sortases of the invention. Such AI sortases will facilitate oligomeric or hyperoligomeric formation of the AI surface protein subunits.

AI Sortases of the invention will typically have a signal peptide sequence within the first 70 amino acid residues. They may also include a transmembrane sequence within 50 amino acid residues of the C terminus. The sortases may also include at least one basic amino acid residue within the last 8 amino acids. Preferably, the sortases have one or more active site residues, such as a catalytic cysteine and histidine.

As shown in Figure 1, AI-1 includes the surface exposed proteins of GBS 80, GBS 52 and GBS 104 and the sortases SAG0647 and SAG0648. AI-1 typically appears as an insertion into the 3' end of the *trmA* gene.

In addition to the open reading frames encoding the AI-1 proteins, AI-1 may also include a divergently transcribed transcriptional regulator such as *araC* (*i.e.*, the transcriptional regulator is located near or adjacent to the AI protein open reading frames, but it transcribed in the opposite direction). It is believed that *araC* may regulate the expression of the AI operon. (*See* Korbel et al., Nature Biotechnology (2004) 22(7): 911 – 917 for a discussion of divergently transcribed regulators in *E. coli*).

AI-1 may also include a sequence encoding a *rho* independent transcriptional terminator (see hairpin structure in Figure 1). The presence of this structure within the adhesin island is thought to interrupt transcription after the GBS 80 open reading frame, leading to increased expression of this surface protein.

A schematic identifying AI-1 within several GBS serotypes is depicted in Figure 2. AI-1 sequences were identified in GBS serotype V, strain isolate 2603; GBS serotype III, strain isolate NEM316; GBS serotype II, strain isolate 18RS21; GBS serotype V, strain isolate CJB111; GBS serotype III, strain isolate COH1 and GBS serotype 1a, strain isolate A909. (Percentages shown are amino acid identity to the 2603 sequence). (An AI-1 was not identified in GBS serotype 1b, strain isolate H36B or GBS serotype 1a, strain isolate 515).

An alignment of AI-1 polynucleotide sequences from serotype V, strain isolates 2603 and CJB111; serotype II, strain isolate 18RS21; serotype III, strain isolates COH1 and NEM316; and serotype 1a, strain isolate A909 is presented in Figure 18. An alignment of amino acid sequences of AI-1 surface protein GBS 80 from serotype V, strain isolates 2603 and CJB111; serotype 1a, strain isolate A909; serotype III, strain isolates COH1 and NEM316 is presented in Figure 22. An alignment of amino acid sequences of AI-1 surface protein GBS 104 from serotype V, strain isolates 2603 and CJB111; serotype III, strain isolates COH1 and NEM316; and serotype II, strain isolate 18RS21 is presented in Figure 23. Preferred AI-1 polynucleotide and amino acid sequences are conserved among two or more GBS serotypes or strain isolates.

As shown in this figure, the full length of surface protein GBS 80 is particularly conserved among GBS serotypes V (strain isolates 2603 and CJBIII), III (strain isolates NEM316 and COH1), and Ia (strain isolate A909). The GBS 80 surface protein is missing or fragmented in serotypes II (strain isolate 18RS21), Ib (strain isolate H36B) and Ia (strain isolate 515).

Polynucleotide and amino acid sequences for *AraC* are set forth in FIGURE 30.

GBS Adhesin Island 2

A second adhesin island, "Adhesin Island 2" or "AI-2" or "GBS AI-2" has also been identified in numerous GBS serotypes. A schematic depicting the correlation between AI-1 and AI-2 within the GBS serotype V, strain isolate 2603 is shown in Figure 3. (Homology percentages in Figure 3 represent amino acid identity of the AI-2 proteins to the AI-1 proteins). Alignments of AI-2 polynucleotide sequences are presented in Figures 20 and 21 (Figure 20 includes sequences from serotype V, strain isolate 2603 and serotype III, strain isolate NEM316. Figure 21 includes sequences from serotype III, strain isolate COH1 and serotype Ia, strain isolate A909). An alignment of amino acid sequences of AI-2 surface protein GBS 067 from serotype V, strain isolates 2603 and CJB111; serotype 1a, strain isolate 515; serotype II, strain isolate 18RS21; serotype Ib, strain isolate H36B; and serotype III, strain isolate NEM316 is presented in Figure 24. Preferred AI-2 polynucleotide and amino acid sequences are conserved among two or more GBS serotypes or strain isolates.

AI-2 comprises a series of approximately five open reading frames encoding for a collection of amino acid sequences comprising surface proteins and sortases. Specifically, AI-2 includes open reading frames encoding for two or more (i.e., 2, 3, 4, 5 or more) of GBS 67, GBS 59, GBS 150, SAG1405, SAG1406, 01520, 01521, 01522, 01523, 01523, 01524 and 01525. In one embodiment, AI-2 includes open reading frames encoding for two or more of GBS 67, GBS 59, GBS 150, SAG1405, and SAG1406. Alternatively, AI-2 may include open reading frames encoding for two or more of 01520, 01521, 01522, 01523, 01523, 01524 and 01525.

One or more of the surface proteins typically include an LPXTG motif (such as LPXTG (SEQ ID NO: 122)) or other sortase substrate motif. The GBS AI-2 sortase proteins are thought to be involved in the secretion and anchoring of the LPXTG containing surface proteins. GBS AI-2 may encode for at least one surface protein. Alternatively, AI-2 may encode for at least two surface proteins and at least one sortase. Preferably, GBS AI-2 encodes for at least three surface proteins and at least two sortases. One or more of the AI-2 surface proteins may include an LPXTG or other sortase substrate motif.

One or more of the surface proteins may also typically include pilin motif. The pilin motif may be involved in pili formation. Cleavage of AI surface proteins by sortase between the threonine and glycine residue of an LPXTG motif yields a thioester-linked acyl intermediate of sortase. The first lysine residue in a pilin motif can serve as an amino group acceptor of the cleaved LPXTG motif and thereby provide a covalent linkage between AI subunits to form pili. For example, the pilin motif can make a nucleophilic attack on the acyl enzyme providing a covalent linkage between AI subunits to form pili and regenerate the sortase enzyme. Some examples of pilin motifs that may be present in the GBS AI-2 proteins include ((YPKN(X_8)K; SEQ ID NO: 158), (PK(X_8)K; SEQ ID NO: 159), (YPK(X_9)K;SEQ ID NO: 160), (PKN(X_8)K; SEQ ID NO: 161), or (PK(X_{10})K; SEQ ID NO: 162)).

One or more of the surface protein may also include an E box motif. The E box motif contains a conserved glutamic acid residue that is believed to be necessary for pilus formation. Some examples of E box motifs may include the amino acid sequences YxLxETxAPxG (SEQ ID NO: 163), YxxxExxAxxGY (SEQ ID NO: 164), YxLxExxxPxDY (SEQ ID NO: 165), or YxLxETxAPxGY (SEQ ID NO: 152).

As shown in Figure 3, GBS AI-2 may include the surface exposed proteins of GBS 67, GBS 59 and GBS 150 and the sortases of SAG1406 and SAG1405. Alternatively, GBS AI-2 may include the proteins 01521, 01524 and 01525 and sortases 01520 and 01522. GBS 067 and 01524 are preferred AI-2 surface proteins.

AI-2 may also include a divergently transcribed transcriptional regulator such as a RofA like protein (for example *rogB*). As in AI-1, *rogB* is thought to regulate the expression of the AI-2 operon.

A schematic depiction of AI-2 within several GBS serotypes is depicted in Figure 4. (Percentages shown are amino acid identity to the 2603 sequence). While the AI-2 surface proteins GBS 59 and GBS 67 are more variable across GBS serotypes than the corresponding AI-1 surface proteins, AI-2 surface protein GBS 67 appears to be conserved in GBS serotypes where the AI-1 surface proteins are disrupted or missing.

For example, as discussed above and in Figure 2, the AI-1 GBS 80 surface protein is fragmented in GBS serotype II, strain isolate 18RS21. Within AI-2 for this same sequence, as shown in Figure 4, the GBS 67 surface protein has 99% amino acid sequence homology with the corresponding sequence in strain isolate 2603. Similarly, the AI-1 GBS 80 surface protein appears to be missing in GBS serotype Ib, strain isolate H36B and GBS serotype Ia, strain isolate 515. Within AI-2 for these sequences, however, the GBS 67 surface protein has 97 – 99 % amino acid sequence homology with the corresponding sequence in strain isolate 2603. GBS 67 appears to have two allelic variants, which can be divided according to percent homology with strains 2603 and H36B. See figures 237-239.

Unlike for GBS 67, amino acid sequence identity of GBS 59 is variable across different GBS strains. As shown in Figures 63 and 224, GBS 59 of GBS strain isolate 2603 shares 100% amino acid residue homology with GBS strain 18RS21, 62% amino acid sequence homology with GBS strain H36B, 48% amino acid residue homology with GBS strain 515 and GBS strain CJB111, and 47% amino acid residue homology with GBS strain NEM316. The amino acid sequence homologies of the different GBS strains suggest that there are two isoforms of GBS 59. The first isoform appears to include the GBS 59 protein of GBS strains CJB111, NEM316, and 515. The second isoform appears to include the GBS 59 protein of GBS strains 18RS21, 2603, and H36B. (See Figures 63 and 224.)

As expected from the variability in GBS 59 isoforms, antibodies specific for the first GBS 59 isoform detect the first but not the second GBS 59 isoform and antibodies specific for the second GBS 59 isoform detect the second but not the first GBS 59 isoform. See Figure 226A, which shows FACS analysis of 28 GBS strains having a GBS 59 gene detected using PCR for GBS 59 surface expression. For each of the 28 GBS strains, FACS analysis was performed using either an antibody for GBS 59 isoform 1 (α-cjb111) or GBS 59 isoform 2 (α -2603). Only one of the two antibodies detected GBS 59 surface expression on each GBS strain. As a negative control, GBS strains in which a GBS 59 gene was not detectable by PCR did not have significant GBS 59 surface expression levels. Figure 226B.

Also, GBS 59 is opsonic only against GBS strains expressing a homologous GBS 59 protein. See Figure 225.

In one embodiment, the immunogenic composition of the invention comprises a first and a second isoform of the GBS 59 protein to provide protection across a wide range of GBS serotypes that express polypeptides from a GBS AI-2. The first isoform may be the GBS 59 protein of GBS strain CJB111, NEM316, or 515. The second isoform may be the GBS 59 protein of GBS strain 18RS21, 2603, or H36B.

The gene encoding GBS 59 has been identified in a high number of GBS isolates; the GBS 59 gene was detected in 31 of 40 GBS isolates tested (77.5%). The GBS 59 protein also appears to be present as part of a pilus in whole extracts derived from GBS strains. Figure 64 shows detection of high molecular weight GBS 59 polymers in whole extracts of GBS strains CJB111, 7357B, COH31, D1363C, 5408, 1999, 5364, 5518, and 515 using antiserum raised against GBS 59 of GBS strain CJB111. Figure 65 also shows detection of these high molecular weight GBS 59 polymers in whole extracts of GBS strains D136C, 515, and CJB111 with anti-GBS 59 antiserum. (See also Figure 220 A for detection of GBS 59 high molecular weight polymers in strain 515.) Figure 65 confirms the presence of different isoforms of GBS 59. Antisera raised against two different GBS 59 isoforms results in different patterns of immunoreactivity depending on the GBS strain origin of the whole extract. Figure 65 further shows detection of GBS 59 monomers in purified GBS 59 preparations.

GBS 59 is also highly expressed on the surface of GBS strains. GBS 59 was detected on the surface of GBS strains CJB111, DK1, DK8, Davis, 515, 2986, 5551, 1169, and 7357B by FACS analysis using mouse antiserum raised against GBS 59 of GBS CJB111. FACS analysis did not detect surface expression of GBS 59 in GBS strains SMU071, JM9130013, and COH1, which do not contain a GBS 59 gene. (See Figure 66.) Further confirmation that GBS 59 is expressed on the surface of GBS is detection of GBS 59 by immuno-electron microscopy on the surface of GBS strain 515 bacteria. See Figure 215.

GBS 67 and GBS 150 also appear to be included in high molecular weight structures, or pili. Figure 69 shows that anti-GBS 67 and anti-GBS 150 immunoreact with high molecular weight structures in whole GBS strain 515 extracts. (See also Figure 220 B and C.) It is also notable in Figure 69 that the anti-GBS 59 antisera, raised in a mouse following immunization with GBS 59 of GBS strain 2603, does not cross-hybridize with GBS 59 in GBS strain 515. GBS 59 of GBS stain 515 is of a different isotype than GBS 59 of GBS stain 2603. See Figure 63, which illustrates that the homology of these two GBS 59 polypeptides is 48%, and Figure 65, which confirms that GBS 59 antisera raised against GBS strain 2603 does not cross-hybridize with GBS 59 of GBS strain 515.

Formation of pili containing GBS 150 does not appear to require GBS 67 expression. Figure 70 provides Western blots showing that higher molecular weight structures in GBS strain 515 total extracts immunoreact with anti-GBS 67 and anti-GBS 150 antiserum. In a GBS strain 515 lacking GBS 67 expression, anti-GBS 67 antiserum no longer immunoreacts with polypeptides in total extracts, while anti-GBS 150 antiserum is still able to cross-hybridze with high molecular weight structures.

Likewise, formation of pili containing GBS 59 does not appear to require GBS 67 expression. As expected, FACS detects GBS 67 cell surface expression on wildtype GBS strain 515, but not GBS strain 515 cells knocked out for GBS 67. FACS analysis using anti-GBS 59 antisera, however, detects GBS 59 expression on both the wildtype GBS strain 515 cells and the GBS strain 515 cells knocked out for GBS 67. Thus, GBS 59 cell surface expression is detected on GBS stain 515 cells regardless of GBS 67 expression.

GBS 67, while present in pili, appears to be localized around the surface of GBS strain 515 cells. See the immuno-electron micrographs presented in Figure 216. GBS 67 binds to fibronectin. See Figure 217.

Formation of pili encoded by GBS AI-2 does require expression of GBS 59. Deletion of GBS 59 from strain 515 bacteria eliminates detection of high molecular weight structures by antibodies that bind to GBS 59 (Figure 221 A, lane 3), GBS 67 (Figure 221 B, lane 3), and GBS 150 (Figure 221 C, lane 3). By contrast, Western blot analysis of 515 bacteria with a deletion of the GBS 67 gene detects high molecular weight structures using GBS 59 (Figure 221 A, lane 2) and GBS 150 (Figure 221 C, lane 2) antisera. Similarly, Western blot analysis of 515 bacteria with a deletion of the GBS 150 gene detects high molecular weight structures using GBS 59 (Figure 221 A, lane 4) and GBS 67 (Figure 221 B, lane 4). See also Figure 223, which provides Western blots of each of the 515 strains interrogated with antibodies for GBS 59, GBS 67, and GBS 150. FACS analysis of strain 515 bacteria deleted for either GBS 59 or GBS 67 confirms these results. See Figure 222, which shows that only deletion of GBS 59 abolishes surface expression of both GBS 59 and GBS 67.

Formation of pili encoded by GBS AI-2 also requires expression of both GBS adhesin island-2 encoded sortases. See Figure 218, which provides Western blot analysis of strain 515 bacteria lacking Srt1, Srt2, or both Srt1 and Srt2. Only deletion of both Srt1 and Srt2 abolishes pilus assembly as detected by antibodies that cross-hybridize with each of GBS 59, GBS 67 and GBS 150. The results of the Western blot analysis were verified by FACS, which provided similar results. See Figure 219.

As shown in Figure 4, two of the GBS strain isolates (COH 1 and A909) do not appear to contain homologues to the surface proteins GBS 59 and GBS 67. For these two strains, the percentages shown in Figure 4 are amino acid identity to the COH1 protein). Notwithstanding the difference in the surface protein lengths for these two strains, AI-2 within these sequences still contains two sortase proteins and three LPXTG containing surface proteins, as well as a signal peptidase sequence leading into the first surface protein. One of the surface proteins in this variant of AI-2, spb1, has previously been identified as a potential adhesion protein. (See Adderson et al., Infection and Immunity (2003) 71(12):6857 – 6863). Alternatively, because of the lack of GBS 59 and GBS 67 sequences, this variant of AI-2 may be a third type of AI (Adhesin Island-3, AI-3, or GBS AI-3).

More than one AI surface protein may be present in the oligomeric, pilus-like structures of the invention. For example, GBS 59 and GBS 67 may be incorporated into an oligomeric structure. Alternatively, GBS 59 and GBS 150 may be incorporated into an oligomeric structure, or GBS 59, GBS 150 and GBS 67 may be incorporated into an oligomeric structure.

In another embodiment, the invention includes compositions comprising two or more AI surface proteins. The composition may include surface proteins from the same adhesin island. For example, the composition may include two or more GBS AI-2 surface proteins, such as GBS 59, GBS 67 and GBS 150. The surface proteins may be isolated from Gram positve bacteria or they may be produced recombinantly.

GAS Adhesin Islands

As discussed above, Applicants have identified at least four different GAS Adhesin Islands. These adhesion islands are thought to encode surface proteins which are important in the bacteria's virulence, and Applicants have obtained the first electron micrographs revealing the presence of these adhesin island proteins in hyperoligomeric pilus structures on the surface of Group A Streptococcus.

Group A Streptococcus is a human specific pathogen which causes a wide variety of diseases ranging from pharyngitis and impetigo through life threatening invasive disease and necrotizing fascilitis. In addition, post-streptococcal autoimmune responses are still a major cause of cardiac pathology in children.

Group A Streptococcal infection of its human host can generally occur in three phases. The first phase involves attachment and/or invasion of the bacteria into host tissue and multiplication of the bacteria within the extracellular spaces. Generally this attachment phase begins in the throat or the skin. The deeper the tissue level infected, the more severe the damage that can be caused. In the second stage of infection, the bacteria secretes a soluble toxin that diffuses into the surrounding tissue or even systemically through the vasculature. This toxin binds to susceptible host cell receptors and triggers innappropropriate immune responses by these host cells, resulting in pathology. Because the toxin can diffuse throughout the host, the necrosis directly caused by the GAS toxins may be physically located in sites distant from the bacterial infection. The final phase of GAS infection can occur long after the original bacteria have been cleared from the host system. At this stage, the host's previous immune response to the GAS bacteria due to cross reactivity between epitopes of a GAS surface protein, M, and host tissues, such as the heart. A general review of GAS infection can be found in Principles of Bacterial Pathogeneis, Groisman ed., Chapter 15 (2001).

In order to prevent the pathogenic effects associated with the later stages of GAS infection, an effective vaccine against GAS will preferably facilitate host elimination of the bacteria during the initial attachment and invasion stage.

Isolates of Group A Streptococcus are historically classified according to the M surface protein described above. The M protein is surface exposed trypsin-sensitive protein generally comprising two polypeptide chains complexed in an alpha helical formation. The carboxyl terminus is anchored in the cytoplasmic membrane and is highly conserved among all group A streptococci. The amino terminus, which extend through the cell wall to the cell surface, is responsible for the antigenic variability observed among the 80 or more serotypes of M proteins.

A second layer of classification is based on a variable, trypsin-resistant surface antigen, commonly referred to as the T-antigen. Decades of epidemiology based on M and T serological typing have been central to studies on the biological diversity and disease causing potential of Group A Streptococci. While the M-protein component and its inherent variability have been extensively characterized, even after five decades of study, there is still very little known about the structure and variability of T-antigens. Antisera to define T types is commercially available from several sources, including Sevapharma (http://www.sevapharma.cz/en).

The gene coding for one form of T-antigen, T-type 6, from an M6 strain of GAS (D741) has been cloned and characterized and maps to an approximately 11 kb highly variable pathogenicity island. Schneewind et al., J Bacteriol. (1990) 172(6):3310 – 3317. This island is known as the Fibronectin-binding, Collagen-binding T-antigen (FCT) region because it contains, in addition to the T6 coding gene (*tee6*), members of a family of genes coding for Extra Cellular Matrix (ECM) binding proteins. Bessen et al., Infection & Immunity (2002) 70(3):1159-1167. Several of the protein products of this gene family have been shown to directly bind either fibronectin and/or collagen. *See* Hanski et al., Infection & Immunity (1992) 60(12):5119-5125; Talay et al., Infection & Immunity (1992(60(9):3837-3844; Jaffe et al. (1996) 21(2):373-384; Rocha et al., Adv Exp Med Biol. (1997) 418:737-739; Kreikemeyer et al., J Biol Chem (2004) 279(16):15850-15859; Podbielski et al., Mol. Microbiol. (1999) 31(4):1051-64; and Kreikemeyer et al., Int. J. Med Microbiol (2004) 294(2-3):177-88. In some cases direct evidence for a role of these proteins in adhesion and invasion has been obtained.

Applicants raised antiserum against a recombinant product of the *tee6* gene and used it to explore the expression of T6 in M6 strain 2724. In immunoblot of mutanolysin extracts of this strain, the antiserum recognized, in addition to a band corresponding to the predicted molecular mass of the product, very high molecular weight ladders ranging in mobility from about 100 kDa to beyond the resolution of the 3-8% gradient gels used.

This pattern of high molecular weight products is similar to that observed in immunoblots of the protein components of the pili identified in *Streptococcus agalactiae* (described above) and previously in *Corynebacterium diphtheriae*. Electron microscropy of strain M6_2724 with antisera specific for the product of *tee6* revealed abundant surface staining and long pilus like structures extending up to 700 nanometers from the bacterial surface, revealing that the T6 protein, one of the antigens recognized in the original Lancefiled serotyping system, is located within a GAS Adhesin Island (GAS AI-1) and forms long covalently linked pilus structures.

Applicants have identified at least four different Group A Streptococcus Adhesin Islands. While these GAS AI sequences can be identified in numerous M types, Applicants have surprisingly discovered a correlation between the four main pilus subunits from the four different GAS AI types and specific T classifications. While other trypsin-resistant surface exposed proteins are likely also implicated in the T classification designations, the discovery of the role of the GAS adhesin islands (and the associated hyper-oligomeric pilus like structures) in T classification and GAS serotype variance has important implications for prevention and treatment of GAS infections. Applicants have identified protein components within each of the GAS adhesin islands which are associated with the pilus formation. These proteins are believed to be involved in the bacteria's initial adherence mechanisms. Immunological recognition of these proteins may allow the host immune response to slow or prevent the bacteria's transition into the more pathogenic later stages of infection.

In addition, Applicants have discovered that the GBS pili structures appear to be implicated in the formation of biofilms (populations of bacteria growing on a surface, often enclosed in an exopolysaccharide matrix). Biofilms are generally associated with bacterial resistance, as antibiotic treatments and host immune response are frequently unable to erradicate all of the bacteria components of the biofilm. Direction of a host immune response against surface proteins exposed during the first steps of bacterial attachment (i.e., before complete biofilm formation) is preferable.

The invention therefore provides for improved immunogenic compositions against GAS infection which may target GAS bacteria during their initial attachment efforts to the host epithelial cells and may provide protection against a wide range of GAS serotypes. The immunogenic compositions of the invention include GAS AI surface proteins which may be formulated in an oligomeric, or hyperoligomeric (pilus) form. The invention also includes combinations of GAS AI surface proteins. Combinations of GAS AI surface proteins may be selected from the same adhesin island or they may be selected from different GAS adhesin islands.

While there is surprising variability in the number and sequence of the GAS AI components across isolates, GAS AI sequences may be generally characterized as Type 1, Type 2, Type 3, and Type 4, depending on the number and type of sortase sequence within the island and the percentage identity of other proteins within the island. Schematics of the GAS adhesin islands are set forth in FIGURE 51A and FIGURE 162. In all strains identified so far, the adhesin island region is flanked by highly conserved open reading frames M1_123 and M1_136. Between three and five genes in each GAS adhesin island code for ECM binding adhesin proteins containing LPXTG motifs.

<u>GAS Adhesin Island 1</u>

As discussed above, Applicants have identified adhesin islands, "GAS Adhesin Island 1" or "GAS AI-1", within the genome Group A Streptococcus serotypes and isolates. GAS AI-1 comprises a series of approximately five open reading frames encoding for a collection of amino acid sequences comprising surface proteins and sortases ("GAS AI-1 proteins"). GAS AI-1 preferably comprises surface proteins, a srtB sortase, and a rofA divergently transcribed transcriptional regulator. GAS AI-1 surface proteins may include a fibronectin binding protein, a collagen adhesion protein and a fimbrial structural subunit. Preferably, each of these GAS AI-1 surface proteins includes an LPXTG sortase substrate motif, such as LPXTG (SEQ ID NO: 122) or LPXSG (SEQ ID NO: 134) (conservative replacement of threonine with serine). Specifically, GAS AI-1 includes open reading frames encoding for two or more (*i.e.*, 2, 3, 4 or 5) of M6_Spy0157, M6_Spy0158, M6_Spy0159, M6_Spy0160, M6_Spy0161.

Applicants have also identified open reading frames encoding fimbrial structural subunits in other GAS bacteria harbouring an AI-1. These open reading frames encode fimbrial structural subunits CDC SS 410_fimbrial, ISS3650_fimbrial, and DSM2071_fimbrial. A GAS AI-1 may comprise a polynucleotide encoding any one of CDC SS 410_fimbrial, ISS3650_fimbrial, and DSM2071_fimbrial.

As discussed above, the hyper-oligomeric pilus structure of GAS AI-1 appears to be responsible for the T-antigen type 6 classification, and GAS AI-1 corresponds to the FCT region previously identified for *tee6*. As in GAS AI-1, the *tee6* FCT region includes open reading frames encoding for a collagen adhesion protein (cpa, capsular polysaccharide adhesion) and a fibronectin binding protein (prtF1). Immunoblots of *tee6*, a GAS AI-1 fimbrial structural subunit corresponding to M6_Spy160, reveal high molecular weight structures indicative of the hyper-oligomeric pilus structures. Immunoblots with antiserum specific for Cpa also recognize a high molecular weight ladder structure, indicating Cpa involvement in the GAS AI-1 pilus structure or formation. In EM photos of GAS bacteria, Cpa antiserum reveals abundant staining on the surface of the bacteria and occasional gold particles extended from the surface of the bacteria. In contrast, immunoblots with antiserum specific for PrtF1 recognize only a single molecular species with electrophoretic mobility corresponding to its predicted molecular mass, indicating that PrtF1 may not be associated with the oligomeric pilus structure. A preferred immunogenic composition of the invention comprises a GAS AI-1 surface protein which may be formulated or purified in an oligomeric (pilis) form. In a preferred embodiment, the oligomeric form is a hyperoligomer. Another preferred immunogenic composition of the invention comprises a GAS AI-1 surface protein which has been isolated in an oligomeric (pilis) form. The oligomer or hyperoligomeric pilus structures comprising the GAS AI-1 surface proteins may be purified or otherwise formulate for use in immunogenic compositions.

One or more of the GAS AI-1 open reading frame polynucleotide sequences may be replaced by a polynucleotide sequence coding for a fragment of the replaced ORF. Alternatively, one or more of the GAS AI-1 open reading frames may be replaced by a sequence having sequence homology to the replaced ORF.

One or more of the GAS AI-1 surface protein sequences typically include an LPXTG motif (such as LPXTG (SEQ ID NO: 122)) or other sortase substrate motif.

The LPXTG sortase substrate motif of a GAS AI surface protein may be generally represented by the formula XXXXG, wherein X at amino acid position 1 is an L, a V, an E, or a Q, wherein X at amino acid position 2 is a P if X at amino acid position 1 is an L, wherein X at amino acid position 2 is a V if X at amino acid position 1 is a E or a Q, wherein X at amino acid position 2 is a V or a P if X at amino acid position 1 is a V, wherein X at amino acid position 3 is any amino acid residue, wherein X at amino acid position 4 is a T if X at amino acid position 1 is a V, E, or Q, and wherein X at amino acid position 4 is a T, S, or A if X at amino acid position 1 is an L. Some examples of LPXTG motifs present in GAS AI surface proteins include LPSXG (SEQ ID NO: 134), VVXTG (SEQ ID NO: 135), EVXTG (SEQ ID NO: 136), VPXTG (SEQ ID NO: 137), QVXTG (SEQ ID NO: 138), LPXAG (SEQ ID NO: 139), QVPTG (SEQ ID NO: 140), and FPXTG (SEQ ID NO: 141).

The GAS AI surface proteins of the invention may affect the ability of the GAS bacteria to adhere to and invade epithelial cells. AI surface proteins may also affect the ability of GAS to translocate through an epithelial cell layer. Preferably, one or more GAS AI surface proteins are capable of binding to or otherwise associating with an epithelial cell surface. GAS AI surface proteins may also be able to bind to or associate with fibrinogen, fibronectin, or collagen.

The GAS AI-1 sortase proteins are predicted to be involved in the secretion and anchoring of the LPXTG containing surface proteins. GAS AI-1 may encode for at least one surface protein. Alternatively, GAS AI-1 may encode for at least two surface exposed proteins and at least one sortase. Preferably, GAS AI-1 encodes for at least three surface exposed proteins and at least two sortases.

The AI surface proteins may be covalently attached to the bacterial cell wall by membrane-associated transpeptidases, such as an AI sortase. The sortase may function to cleave the surface protein, preferably between the threonine and glycine residues of an LPXTG motif. The sortase may then assist in the formation of an amide link between the threonine carboxyl group and a cell wall precursor such as lipid II. The precursor can then be incorporated into the peptidoglycan via the transglycoslylation and transpeptidation reactions of bacterial wall synthesis. *See* Comfort et al., Infection & Immunity (2004) 72(5): 2710 – 2722.

GAS AI-1 preferably includes a srtB sortase. GAS srtB sortases may preferably anchor surface proteins with an LPSTG motif (SEQ ID NO: 166), particularly where the motif is followed by a serine.

In one embodiment, the invention includes a composition comprising oligomeric, pilus-like structures comprising a GAS AI-1 surface protein such as M6_Spy0157, M6_Spy0159, M6_Spy0160, CDC SS 410_fimbrial, ISS3650_fimbrial, or DSM2071_fimbrial. The oligomeric, pilus-like structure may comprise numerous units of AI surface protein. Preferably, the oligomeric, pilus-like structures comprise two or more AI surface proteins. Still more preferably, the oligomeric, pilus-like structure comprises a hyper-oligomeric pilus-like structure comprising at least two (*e.g.*, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 20, 25, 30, 35, 40, 45, 50, 60, 70, 80, 90, 100, 120, 140, 150, 200 or more) oligomeric subunits, wherein each subunit comprises an AI surface protein or a fragment thereof. The oligomeric subunits may be covalently associated via a conserved lysine within a pilin motif. The oligomeric subunits may be covalently associated via an LPXTG motif, preferably, via the threonine or serine amino acid residue, respectively.

AI surface proteins or fragments thereof to be incorporated into the oligomeric, pilus-like structures of the invention will preferably include a pilin motif.

The oligomeric, pilus like structures may be used alone or in the combinations of the invention. In one embodiment, the invention comprises a GAS Adhesin Island protein in oligomeric form, preferably in a hyperoligomeric form. In one embodiment, the invention comprises a composition comprising one or more GAS Adhesin Island 1 ("GAS AI-1") proteins and one or more GAS Adhesin Island 2 ("GAS AI-2"), GAS Adhesin Island 3 ("GAS AI-3"), or GAS Adhesin Island 4 ("GAS AI-4") proteins, wherein one or more of the GAS Adhesin Island proteins is in the form of an oligomer, preferably in a hyperoligomeric form.

In addition to the open reading frames encoding the GAS AI-1 proteins, GAS AI-1 may also include a divergently transcribed transcriptional regulator such as *RofA* (*i.e.*, the transcriptional regulator is located near or adjacent to the AI protein open reading frames, but it transcribed in the opposite direction).

GAS Adhesin Island 2

A second adhesin island, "GAS Adhesin Island 2" or "GAS AI-2" has also been identified in Group A Streptococcus serotypes and isolates. GAS AI-2 comprises a series of approximately eight open reading frames encoding for a collection of amino acid sequences comprising surface proteins and sortases ("GAS AI-2 proteins"). Specifically, GAS AI-2 includes open reading frames encoding for two or more (*i.e.*, 2, 3, 4, 5, 6, 7, or 8) of GAS15, Spy0127, GAS16, GAS17, GAS18, Spy0131, Spy0133, and GAS20.

A preferred immunogenic composition of the invention comprises a GAS AI-2 surface protein which may be formulated or purified in an oligomeric (pilis) form. In a preferred embodiment, the oligomeric form is a hyperoligomer. Another preferred immunogenic composition of the invention comprises a GAS AI-2 surface protein which has been isolated in an oligomeric (pilis) form. The oligomer or hyperoligomeric pilus structures comprising the GAS AI-2 surface proteins may be purified or otherwise formulate for use in immunogenic compositions.

One or more of the GAS AI-2 open reading frame polynucleotide sequences may be replaced by a polynucleotide sequence coding for a fragment of the replaced ORF. Alternatively, one or more of the GAS AI-2 open reading frames may be replaced by a sequence having sequence homology to the replaced ORF.

One or more of the GAS AI-2 surface protein sequences typically include an LPXTG motif (such as LPXTG (SEQ ID NO: 122)) or other sortase substrate motif. The AI surface proteins of the invention may affect the ability of the GAS bacteria to adhere to and invade epithelial cells. AI surface proteins may also affect the ability of GAS to translocate through an epithelial cell layer. Preferably, one or more AI surface proteins are capable of binding to or otherwise associating with an epithelial cell surface. AI surface proteins may also be able to bind to or associate with fibrinogen, fibronectin, or collagen.

The GAS AI-2 sortase proteins are predicted to be involved in the secretion and anchoring of the LPXTG containing surface proteins. GAS AI-2 may encode for at least one surface protein. Alternatively, GAS AI-2 may encode for at least two surface exposed proteins and at least one sortase. Preferably, GAS AI-2 encodes for at least three surface exposed proteins and at least two sortases.

The AI surface proteins may be covalently attached to the bacterial cell wall by membrane-associated transpeptidases, such as an AI sortase. The sortase may function to cleave the surface protein, preferably between the threonine and glycine residues of an LPXTG motif. The sortase may then assist in the formation of an amide link between the threonine carboxyl group and a cell wall precursor such as lipid II. The precursor can then be incorporated into the peptidoglycan via the transglycoslylation and transpeptidation reactions of bacterial wall synthesis. *See* Comfort et al., Infection & Immunity (2004) 72(5): 2710 – 2722.

In one embodiment, the invention includes a composition comprising oligomeric, pilus-like structures comprising an AI surface protein such as GAS15, GAS16, or GAS18. The oligomeric, pilus-like structure may comprise numerous units of AI surface protein. Preferably, the oligomeric, pilus-like structures comprise two or more AI surface proteins. Still more preferably, the oligomeric, pilus-like structure comprises a hyper-oligomeric pilus-like structure comprising at least two (*e.g.*, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 20, 25, 30, 35, 40, 45, 50, 60, 70, 80, 90, 100, 120, 140, 150, 200 or more) oligomeric subunits, wherein each subunit comprises an AI surface protein or a fragment thereof. The oligomeric subunits may be covalently associated via a conserved lysine within a pilin motif. The oligomeric subunits may be covalently associated via an LPXTG motif, preferably, via the threonine amino acid residue.

AI surface proteins or fragments thereof to be incorporated into the oligomeric, pilus-like structures of the invention will preferably include a pilin motif.

The oligomeric, pilus like structures may be used alone or in the combinations of the invention. In one embodiment, the invention comprises a GAS Adhesin Island protein in oligomeric form, preferably in a hyperoligomeric form. In one embodiment, the invention comprises a composition comprising one or more GAS Adhesin Island 2 ("GAS AI-2") proteins and one or more GAS Adhesin Island 1 ("GAS AI-1"), GAS Adhesin Island 3 ("GAS AI-3"), or GAS Adhesin Island 4 ("GAS AI-4") proteins, wherein one or more of the Adhesin Island proteins is in the form of an oligomer, preferably in a hyperoligomeric form.

In addition to the open reading frames encoding the GAS AI-2 proteins, GAS AI-2 may also include a divergently transcribed transcriptional regulator such as *rofA* (*i.e.*, the transcriptional regulator is located near or adjacent to the AI protein open reading frames, but it transcribed in the opposite direction).

GAS Adhesin Island 3

A third adhesin island, "GAS Adhesin Island 3" or "GAS AI-3" has also been identified in several Group A Streptococcus serotypes and isolates. GAS AI-3 comprises a series of approximately seven open reading frames encoding for a collection of amino acid sequences comprising surface proteins and sortases ("GAS AI-3 proteins"). Specifically, GAS AI-3 includes open reading frames encoding for two or more (*i.e.*, 2, 3, 4, 5, 6, or 7) of SpyM3_0098, SpyM3_0099, SpyM3_0100, SpyM3_0101, SpyM3_0102, SpyM3_0103, SpyM3_0104, SPs0100, SPs0101, SPs0102, SPs0103, SPs0104, SPs0105, SPs0106, orf78, orf79, orf80, orf81, orf82, orf83, orf84, spyM18_0126, spyM18_0127, spyM18_0128, spyM18_0129, spyM18_0130, spyM18_0131, spyM18_0132, SpyoM01000156, SpyoM01000155, SpyoM01000154, SpyoM01000153, SpyoM01000152, SpyoM01000151, SpyoM01000150, and SpyoM01000149. In one embodiment, GAS AI-3 includes open reading frames encoding for two or more (*i.e.*, 2, 3, 4, 5, 6, or 7) of SpyM3_0098, SpyM3_0099, SpyM3_0100, SpyM3_0101, SpyM3_0102, SpyM3_0103, and SpyM3_0104. In another embodiment, GAS AI-3 includes open reading frames encoding for two or more (*i.e.*, 2, 3, 4, 5, 6, or 7) of SPs0100, SPs0101, SPs0102, SPs0103, SPs0104, SPs0105, and SPs0106. In a further embodiment, GAS AI-3 includes open reading frames encoding for two or more (*i.e.*, 2, 3, 4, 5, 6, or 7) of orf78, orf79, orf80, orf81, orf82, orf83, and orf84. In yet another embodiment, GAS AI-3 includes open reading frames encoding for two or more (*i.e.*, 2, 3, 4, 5, 6, or 7) of spyM18_0126, spyM18_0127, spyM18_0128, spyM18_0129, spyM18_0130, spyM18_0131, and spyM18_0132. In yet another embodiment, GAS AI-3 includes open reading frames encoding for two or more (*i.e.*, 2, 3, 4, 5, 6, or 7) of SpyoM01000156, SpyoM01000155, SpyoM01000154, SpyoM01000153, SpyoM01000152, SpyoM01000151, SpyoM01000150, and SpyoM01000149.

Applicants have also identified open reading frames encoding fimbrial structural subunits in other GAS bacteria harbouring an AI-3. These open reading frames encode fimbrial structural subunits ISS3040_fimbrial, ISS3776_fimbrial, and ISS4959_fimbrial. A GAS AI-3 may comprise a polynucleotide encoding any one of ISS3040_fimbrial, ISS3776_fimbrial, and ISS4959_fimbrial.

One or more of the GAS AI-3 open reading frame polynucleotide sequences may be replaced by a polynucleotide sequence coding for a fragment of the replaced ORF. Alternatively, one or more of the GAS AI-3 open reading frames may be replaced by a sequence having sequence homology to the replaced ORF.

A preferred immunogenic composition of the invention comprises a GAS AI-3 surface protein which may be formulated or purified in an oligomeric (pilis) form. In a preferred embodiment, the oligomeric form is a hyperoligomer. Another preferred immunogenic composition of the invention comprises a GAS AI-3 surface protein which has been isolated in an oligomeric (pilis) form. The oligomer or hyperoligomeric pilus structures comprising the GAS AI-3 surface proteins may be purified or otherwise formulate for use in immunogenic compositions.

One or more of the GAS AI-3 surface protein sequences typically include an LPXTG motif (such as LPXTG (SEQ ID NO: 122)) or other sortase substrate motif. The AI surface proteins of the invention may affect the ability of the GAS bacteria to adhere to and invade epithelial cells. AI surface proteins may also affect the ability of GAS to translocate through an epithelial cell layer. Preferably, one or more AI surface proteins are capable of binding to or otherwise associating with an epithelial cell surface. AI surface proteins may also be able to bind to or associate with fibrinogen, fibronectin, or collagen.

The GAS AI-3 sortase proteins are predicted to be involved in the secretion and anchoring of the LPXTG containing surface proteins. GAS AI-3 may encode for at least one surface protein. Alternatively, GAS AI-3 may encode for at least two surface exposed proteins and at least one sortase. Preferably, GAS AI-3 encodes for at least three surface exposed proteins and at least two sortases.

The AI surface proteins may be covalently attached to the bacterial cell wall by membrane-associated transpeptidases, such as an AI sortase. The sortase may function to cleave the surface protein, preferably between the threonine and glycine residues of an LPXTG motif. The sortase may then assist in the formation of an amide link between the threonine or alanine carboxyl group and a cell wall precursor such as lipid II. The precursor can then be incorporated into the peptidoglycan via the transglycoslylation and transpeptidation reactions of bacterial wall synthesis. *See* Comfort et al., Infection & Immunity (2004) 72(5): 2710 – 2722.

The invention includes a composition comprising oligomeric, pilus-like structures comprising an AI surface protein such as SpyM3_0098, SpyM3_0100, SpyM3_0102, SpyM3_0104, SPs0100, SPs0102, SPs0104, SPs0106, orf78, orf80, orf82, orf84, spyM18_0126, spyM18_0128, spyM18_0130, spyM18_0132, SpyoM01000155, SpyoM01000153, SpyoM01000151, SpyoM01000149, ISS3040_fimbrial, ISS3776_fimbrial, and ISS4959_fimbrial. In one embodiment, the invention includes a composition comprising oligomeric, pilus-like structures comprising an AI surface protein such as SpyM3_0098, SpyM3_0100, SpyM3_0102, and SpyM3_0104. In another embodiment, the invention includes a composition comprising oligomeric, pilus-like structures comprising an AI surface protein such as SPs0100, SPs0102, SPs0104, and SPs0106. In another embodiment, the invention includes a composition comprising oligomeric, pilus-like structures comprising an AI surface protein such as orf78, orf80, orf82, and orf84. In yet another embodiment, the invention includes a composition comprising oligomeric, pilus-like structures comprising an AI surface protein such as spyM18_0126, spyM18_0128, spyM18_0130, and spyM18_0132. In a further embodiment, the invention includes a composition comprising oligomeric, pilus-like structures comprising an AI surface protein such as SpyoM01000155, SpyoM01000153, SpyoM01000151, and SpyoM01000149. In yet a further embodiment, the invention includes a composition comprising oligomeric, pilus-like structures comprising an AI surface protein such as ISS3040_fimbrial, ISS3776_fimbrial, and ISS4959_fimbrial. The oligomeric, pilus-like structure may comprise numerous units of AI surface protein. Preferably, the oligomeric, pilus-like structures comprise two or more AI surface proteins. Still more preferably, the oligomeric, pilus-like structure comprises a hyper-oligomeric pilus-like structure comprising at least two (*e.g.*, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 20, 25, 30, 35, 40, 45, 50, 60, 70, 80, 90, 100, 120, 140, 150, 200 or more) oligomeric subunits, wherein each subunit comprises an AI surface protein or a fragment thereof. The oligomeric subunits may be covalently associated via a conserved lysine within a pilin motif. The oligomeric subunits may be covalently associated via an LPXTG motif, preferably, via the threonine amino acid residue.

AI surface proteins or fragments thereof to be incorporated into the oligomeric, pilus-like structures of the invention will preferably include a pilin motif.

The oligomeric, pilus like structures may be used alone or in the combinations of the invention. In one embodiment, the invention comprises a GAS Adhesin Island protein in oligomeric form, preferably in a hyperoligomeric form. In one embodiment, the invention comprises a composition comprising one or more GAS Adhesin Island 3 ("GAS AI-3") proteins and one or more GAS Adhesin Island 1 ("GAS AI-1"), GAS Adhesin Island 2 ("GAS AI-2"), or GAS Adhesin Island 4 ("GAS AI-4") proteins, wherein one or more of the Adhesin Island proteins is in the form of an oligomer, preferably in a hyperoligomeric form.

In addition to the open reading frames encoding the GAS AI-3 proteins, GAS AI-3 may also include a transcriptional regulator such as *Nra*.

GAS Adhesin Island 4

A fourth adhesin island, "GAS Adhesin Island 4" or "GAS AI-4" has also been identified in Group A Streptococcus serotypes and isolates. GAS AI-4 comprises a series of approximately eight open reading frames encoding for a collection of amino acid sequences comprising surface proteins and sortases ("GAS AI-4 proteins"). Specifically, GAS AI-4 includes open reading frames encoding for two or more (*i.e.*, 2, 3, 4, 5, 6, 7, or 8) of 19224134, 19224135, 19223136, 19223137, 19224138, 19224139, 19224140, and 19224141.

Applicants have also identified open reading frames encoding fimbrial structural subunits in other GAS bacteria harbouring an AI-4. These open reading frames encode fimbrial structural subunits 20010296_fimbrial, 20020069_fimbrial, CDC SS 635_fimbrial, ISS4883_fimbrial, and ISS4538_fimbrial. A GAS AI-4 may comprise a polynucleotide encoding any one of 20010296_fimbrial, 20020069_fimbrial, CDC SS 635_fimbrial, ISS4883_fimbrial, and ISS4538_fimbrial.

One or more of the GAS AI-4 open reading frame polynucleotide sequences may be replaced by a polynucleotide sequence coding for a fragment of the replaced ORF. Alternatively, one or more of the GAS AI-4 open reading frames may be replaced by a sequence having sequence homology to the replaced ORF.

A preferred immunogenic composition of the invention comprises a GAS AI-4 surface protein which may be formulated or purified in an oligomeric (pilis) form. In a preferred embodiment, the oligomeric form is a hyperoligomer. Another preferred immunogenic composition of the invention comprises a GAS AI-4 surface protein which has been isolated in an oligomeric (pilis) form. The oligomer or hyperoligomeric pilus structures comprising the GAS AI-4 surface proteins may be purified or otherwise formulate for use in immunogenic compositions.

One or more of the GAS AI-4 surface protein sequences typically include an LPXTG motif (such as LPXTG (SEQ ID NO: 122)) or other sortase substrate motif. The AI surface proteins of the invention may effect the ability of the GAS bacteria to adhere to and invade epithelial cells. AI surface proteins may also affect the ability of GAS to translocate through an epithelial cell layer. Preferably, one or more AI surface proteins are capable of binding to or otherwise associating with an epithelial cell surface. AI surface proteins may also be able to bind to or associate with fibrinogen, fibronectin, or collagen.

The GAS AI-4 sortase proteins are predicted to be involved in the secretion and anchoring of the LPXTG containing surface proteins. GAS AI-4 may encode for at least one surface protein. Alternatively, GAS AI-4 may encode for at least two surface exposed proteins and at least one sortase. Preferably, GAS AI-4 encodes for at least three surface exposed proteins and at least two sortases.

The AI surface proteins may be covalently attached to the bacterial cell wall by membrane-associated transpeptidases, such as an AI sortase. The sortase may function to cleave the surface protein, preferably between the threonine and glycine residues of an LPXTG motif. The sortase may then assist in the formation of an amide link between the threonine carboxyl group and a cell wall precursor such as lipid II. The precursor can then be incorporated into the peptidoglycan via the transglycoslylation and transpeptidation reactions of bacterial wall synthesis. *See* Comfort et al., Infection & Immunity (2004) 72(5): 2710 – 2722.

In one embodiment, the invention includes a composition comprising oligomeric, pilus-like structures comprising an AI surface protein such as 19224134, 19224135, 19224137, 19224139, 19224141, 20010296_fimbrial, 20020069_fimbrial, CDC SS 635_fimbrial, ISS4883_fimbrial, and ISS4538_fimbrial. The oligomeric, pilus-like structure may comprise numerous units of AI surface protein. Preferably, the oligomeric, pilus-like structures comprise two or more AI surface proteins. Still more preferably, the oligomeric, pilus-like structure comprises a hyper-oligomeric pilus-like structure comprising at least two (*e.g.*, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 20, 25, 30, 35, 40, 45, 50, 60, 70, 80, 90, 100, 120, 140, 150, 200 or more) oligomeric subunits, wherein each subunit comprises an AI surface protein or a fragment thereof. The oligomeric subunits may be covalently associated via a conserved lysine within a pilin motif. The oligomeric subunits may be covalently associated via an LPXTG motif, preferably, via the threonine amino acid residue.

AI surface proteins or fragments thereof to be incorporated into the oligomeric, pilus-like structures of the invention will preferably include a pilin motif.

The oligomeric, pilus like structures may be used alone or in the combinations of the invention. In one embodiment, the invention comprises a GAS Adhesin Island protein in oligomeric form, preferably in a hyperoligomeric form. In one embodiment, the invention comprises a composition comprising one or more GAS Adhesin Island 4 ("GAS AI-4") proteins and one or more GAS Adhesin Island 1 ("GAS AI-1"), GAS Adhesin Island 2 ("GAS AI-2"), or GAS Adhesin Island 3 ("GAS AI-3") proteins, wherein one or more of the Adhesin Island proteins is in the form of an oligomer, preferably in a hyperoligomeric form.

In addition to the open reading frames encoding the GAS AI-4 proteins, GAS AI-4 may also include a divergently transcribed transcriptional regulator such as *rofA* (*i.e.*, the transcriptional regulator is located near or adjacent to the AI protein open reading frames, but it transcribed in the opposite direction).

The oligomeric, pilus-like structures of the invention may be combined with one or more additional GAS proteins. In one embodiment, the oligomeric, pilus-like structures comprise one or more AI surface proteins in combination with a second GAS protein.

The oligomeric, pilus-like structures may be isolated or purified from bacterial cultures in which the bacteria express an AI surface protein. The invention therefore includes a method for manufacturing an oligomeric AI surface antigen comprising culturing a GAS bacterium that expresses the oligomeric AI protein and isolating the expressed oligomeric AI protein from the GAS bacteria. The AI protein may be collected from secretions into the supernatant or it may be purified from the bacterial surface. The method may further comprise purification of the expressed AI protein. Preferably, the AI protein is in a hyperoligomeric form.

The oligomeric, pilus-like structures may be isolated or purified from bacterial cultures overexpressing an AI surface protein. The invention therefore includes a method for manufacturing an oligomeric Adhesin Island surface antigen comprising culturing a GAS bacterium adapted for increased AI protein expression and isolation of the expressed oligomeric Adhesin Island protein from the GAS bacteria. The AI protein may be collected from secretions into the supernatant or it may be purified from the bacterial surface. The method may further comprise purification of the expressed Adhesin Island protein. Preferably, the Adhesin Island protein is in a hyperoligomeric form.

The GAS bacteria are preferably adapted to increase AI protein expression by at least two (*e.g.*, 2, 3, 4, 5, 8, 10, 15, 20, 25, 30, 35, 40, 45, 50, 60, 70, 80, 90, 100, 125, 150 or 200) times wild type expression levels.

GAS bacteria may be adapted to increase AI protein expression by any means known in the art, including methods of increasing gene dosage and methods of gene upregulation. Such means include, for example, transformation of the GAS bacteria with a plasmid encoding the AI protein. The plasmid may include a strong promoter or it may include multiple copies of the sequence encoding the AI protein. Optionally, the sequence encoding the AI protein within the GAS bacterial genome may be deleted. Alternatively, or in addition, the promoter regulating the GAS Adhesin Island may be modified to increase expression.

The invention further includes GAS bacteria which have been adapted to produce increased levels of AI surface protein. In particular, the invention includes GAS bacteria which have been adapted to produce oligomeric or hyperoligomeric AI surface protein. In one embodiment, the Gram positive bacteria of the invention are inactivated or attenuated to permit *in vivo* delivery of the whole bacteria, with the AI surface protein exposed on its surface.

The invention further includes GAS bacteria which have been adapted to have increased levels of expressed AI protein incorporated in pili on their surface. The GAS bacteria may be adapted to have increased exposure of oligomeric or hyperoligomeric AI proteins on its surface by increasing expression levels of LepA polypeptide, or an equivalent signal peptidase, in the GAS bacteria. Applicants have shown that deletion of LepA in strain SF370 bacteria, which harbour a GAS AI-2, abolishes surface exposure of M and pili proteins on the GAS. Increased levels of LepA expression in GAS are expected to result in increased exposure of M and pili proteins on the surface of GAS. Increased expression of LepA in GAS may be achieved by any means known in the art, such as increasing gene dosage and methods of gene upregulation. The GAS bacteria adapted to have increased levels of LepA expression may additionally be adapted to express increased levels of at least one pili protein.

Alternatively, the AI proteins of the invention may be expressed on the surface of a non-pathogenic Gram positive bacteria, such as *Streptococus gordonii* (See, e.g., Byrd et al., "Biological consequences of antigen and cytokine co-expression by recombinant *Streptococcus gordonii* vaccine vectors", Vaccine (2002) 20:2197-2205) or *Lactococcus lactis* (See, e.g., Mannam et al., "Mucosal Vaccine Made from Live, Recombinant Lactococcus lactis Protects Mice against Pharangeal Infection with *Streptococcus pyogenes*" Infection and Immunity (2004) 72(6):3444-3450). As used herein, non-pathogenic Gram positive bacteria refer to Gram positive bacteria which are compatible with a human host subject and are not associated with human pathogenisis. Preferably, the non-pathogenic bacteria are modified to express the AI surface protein in oligomeric, or hyper-oligomeric form. Sequences encoding for an AI surface protein and, optionally, an AI sortase, may be integrated into the non-pathogenic Gram positive bacterial genome or inserted into a plasmid. The non-pathogenic Gram positive bacteria may be inactivated or attenuated to facilitate *in vivo* delivery of the whole bacteria, with the AI surface protein exposed on its surface. Alternatively, the AI surface protein may be isolated or purified from a bacterial culture of the non-pathogenic Gram positive bacteria. For example, the AI surface protein may be isolated from cell extracts or culture supernatants. Alternatively, the AI surface protein may be isolated or purified from the surface of the non-pathogenic Gram positive bacteria.

The non-pathogenic Gram positive bacteria may be used to express any of the GAS Adhesin Island proteins described herein. The non-pathogenic Gram positive bacteria are transformed to express an Adhesin Island surface protein. Preferably, the non-pathogenic Gram positive bacteria also express at least one Adhesin Island sortase. The AI transformed non-pathogenic Gram positive bacteria of the invention may be used to prevent or treat infection with pathogenic GAS.

Applicants modified *L. lactis* to demonstrate that, like GBS polypeptides, it can express GAS AI polypeptides. *L. lactis* was transformed with pAM401 constructs encoding entire pili gene clusters of AI-1, AI-2, and AI-4 adhesin islands. Briefly, the pAM401 is a promoterless high-copy plasmid. The entire pili gene clusters of an M6 (AI-1), M1 (AI-2), and M12 (AI-4) bacteria were inserted into the pAM401 construct. The gene clusters were transcribed under the control their own (M6, M1, or M12) promoter or the GBS promoter that successfully initiated expression of the GBS AI-1 adhesin islands in *L. lactis*, described above. Figure 172 provides a schematic depiction of GAS M6 (AI-1), M1 (AI-2), and M12 (AI-4) adhesin islands and indicates the portions of the adhesin island sequences inserted in the pAM401 construct.

Each of the *L. lactis* transformed with one of the M6, M1, or M12 adhesin island gene clusters expressed high molecular weight structures that were immunoreactive with antibodies that bind to polypeptides present in their respective pili. Figures 173 A-C provide results of Western blot analysis of surface protein-enriched extracts of *L. lactis* transformed with M6 (Figure 173 A), M1 (Figure 173 B), or M12 (Figure 173 C) adhesin island gene clusters using antibodies that bind to the fimbrial structural subunit encoded by each cluster. Figure 173A at lanes 3 and 4 shows detection of high molecular structures in *L. lactis* transformed with an adhesin island pilus gene cluster from an M1 AI-2 using an antibody that binds to fimbrial structural subunit Spy0128. Figure 173B at lanes 3 and 4 shows detection of high molecular weight structures in *L. lactis* transformed with an adhesin island pilus gene cluster from an M12 AI-4 using an antibody that binds to fimbrial structural subunit EftLSL.A. Figure 173C at lane 3 shows detection of high molecular weight structures in *L. lactis* transformed with an adhesin island pilus gene cluster from an M6 AI-1 using an antibody that binds to fimbrial structural subunit M6_Spy0160. In figures 173 A-C, "p1" immediately following the notation of AI subtype indicates that the promoter present in the Adhesin Island is used to drive transcription of the adhesin island gene cluster and "p2" indicates that the promoter was the GBS promoter described above. Thus, it appears that *L. lactis* is capable of expressing the fimbrial structural subunits encoded by GAS adhesin islands in an oligomeric form.

Alternatively, the oligomeric, pilus-like structures may be produced recombinantly. If produced in a recombinant host cell system, the AI surface protein will preferably be expressed in coordination with the expression of one or more of the AI sortases of the invention. Such AI sortases will facilitate oligomeric or hyperoligomeric formation of the AI surface protein subunits.

S. pneumoniae from TIGR4 Adhesin Island

As discussed above, Applicants have identified adhesin islands within the genome of *S. pneumoniae* from TIGR4. The *S. pneumoniae* from TIGR4 Adhesin Island comprises a series of approximately seven open reading frames encoding for a collection of amino acid sequences comprising surface proteins and sortases. Specifically, the *S. pneumoniae* from TIGR4 AI proteins includes open reading frames encoding for two or more (*i.e.*, 2, 3, 4, 5, 6, or 7) of SP0462, SP0463, SP0464, SP0465, SP0466, SP0467, and SP0468.

A preferred immunogenic composition of the invention comprises a *S. pneumoniae* from TIGR4 AI surface protein which may be formulated or purified in an oligomeric (pilis) form. In a preferred embodiment, the oligomeric form is a hyperoligomer. Another preferred immunogenic composition of the invention comprises a *S. pneumoniae* from TIGR4 AI surface protein which has been isolated in an oligomeric (pilis) form. The oligomer or hyperoligomer pilus structures comprising *S. pneumoniae* surface proteins may be purified or otherwise formulated for use in immunogenic compositions.

One or more of the *S. pneumoniae* from TIGR4 AI open reading frame polynucleotide sequences may be replaced by a polynucleotide sequence coding for a fragment of the replaced ORF. Alternatively, one or more of the *S. pneumoniae* from TIGR4 AI open reading frames may be replaced by a sequence having sequence homology to the replaced ORF.

One or more of the *S. pneumoniae* from TIGR4 AI surface protein sequences typically include an LPXTG motif (such as LPXTG (SEQ ID NO: 122)) or other sortase substrate motif.

The *S. pneumoniae* from TIGR4 AI surface proteins of the invention may affect the ability of the *S. pneumoniae* bacteria to adhere to and invade epithelial cells. AI surface proteins may also affect the ability of *S. pneumoniae* to translocate through an epithelial cell layer. Preferably, one or more *S. pneumoniae* from TIGR4 AI surface proteins are capable of binding to or otherwise associating with an epithelial cell surface. *S. pneumoniae* from TIGR4 AI surface proteins may also be able to bind to or associate with fibrinogen, fibronectin, or collagen.

The *S. pneumoniae* from TIGR4 AI sortase proteins are predicted to be involved in the secretion and anchoring of the LPXTG containing surface proteins. *S. pneumoniae* from TIGR4 AI may encode for at least one surface protein. Alternatively, *S. pneumoniae* from TIGR4 AI may encode for at least two surface exposed proteins and at least one sortase. Preferably, *S. pneumoniae* from TIGR4 AI encodes for at least three surface exposed proteins and at least two sortases.

The AI surface proteins may be covalently attached to the bacterial cell wall by membrane-associated transpeptidases, such as an AI sortase. The sortase may function to cleave the surface protein, preferably between the threonine and glycine residues of an LPXTG motif. The sortase may then assist in the formation of an amide link between the threonine carboxyl group and a cell wall precursor such as lipid II. The precursor can then be incorporated into the peptidoglycan via the transglycoslylation and transpeptidation reactions of bacterial wall synthesis. *See* Comfort et al., Infection & Immunity (2004) 72(5): 2710 – 2722.

In one embodiment, the invention includes a composition comprising oligomeric, pilus-like structures comprising a *S. pneumoniae* from TIGR4 AI surface protein such as SP0462, SP0463, SP0464, or SP0465. The oligomeric, pilus-like structure may comprise numerous units of AI surface protein. Preferably, the oligomeric, pilus-like structures comprise two or more AI surface proteins. Still more preferably, the oligomeric, pilus-like structure comprises a hyper-oligomeric pilus-like structure comprising at least two (*e.g.*, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 20, 25, 30, 35, 40, 45, 50, 60, 70, 80, 90, 100, 120, 140, 150, 200 or more) oligomeric subunits, wherein each subunit comprises an AI surface protein or a fragment thereof. The oligomeric subunits may be covalently associated via a conserved lysine within a pilin motif. The oligomeric subunits may be covalently associated via an LPXTG motif, preferably, via the threonine or serine amino acid residue, respectively.

AI surface proteins or fragments thereof to be incorporated into the oligomeric, pilus-like structures of the invention will preferably include a pilin motif.

The oligomeric, pilus like structures may be used alone or in the combinations of the invention. In one embodiment, the invention comprises a *S. pneumoniae* from TIGR4 AI protein in oligomeric form, preferably in a hyperoligomeric form. In one embodiment, the invention comprises a composition comprising one or more *S. pneumoniae* from TIGR4 AI proteins and one or more *S. pneumoniae* strain 670 AI proteins, wherein one or more of the *S. pneumoniae* AI proteins is in the form of an oligomer, preferably in a hyperoligomeric form.

In addition to the open reading frames encoding the *S. pneumoniae* from TIGR4 AI proteins, *S. pneumoniae* from TIGR4 AI may also include a transcriptional regulator.

S. pneumoniae strain 670 Adhesin Island

As discussed above, Applicants have identified adhesin islands within the genome of *S. pneumoniae* strain 670. The *S. pneumoniae* strain 670 Adhesin Island comprises a series of approximately seven open reading frames encoding for a collection of amino acid sequences comprising surface proteins and sortases. Specifically, the *S. pneumoniae* strain 670 AI proteins includes open reading frames encoding for two or more (*i.e.*, 2, 3, 4, 5, 6, or 7) of orf1_670, orf3_670, orf4_670, orf5_670, orf6_670, orf7_670, orf8_670.

A preferred immunogenic composition of the invention comprises a *S. pneumoniae* strain 670 AI surface protein which may be formulated or purified in an oligomeric (pilis) form. Another preferred immunogenic composition of the invention comprises a *S. pneumoniae* strain 670 AI surface protein which has been isolated in an oligomeric (pilis) form.

One or more of the *S. pneumoniae* strain 670 AI open reading frame polynucleotide sequences may be replaced by a polynucleotide sequence coding for a fragment of the replaced ORF. Alternatively, one or more of the *S. pneumoniae* strain 670 AI open reading frames may be replaced by a sequence having sequence homology to the replaced ORF.

One or more of the *S. pneumoniae* strain 670 AI surface protein sequences typically include an LPXTG motif (such as LPXTG (SEQ ID NO: 122)) or other sortase substrate motif.

The *S. pneumoniae* strain 670 AI surface proteins of the invention may affect the ability of the *S. pneumoniae* bacteria to adhere to and invade epithelial cells. AI surface proteins may also affect the ability of *S. pneumoniae* to translocate through an epithelial cell layer. Preferably, one or more *S. pneumoniae* strain 670 AI surface proteins are capable of binding to or otherwise associating with an epithelial cell surface. *S. pneumoniae* strain 670 AI surface proteins may also be able to bind to or associate with fibrinogen, fibronectin, or collagen.

The *S. pneumoniae* strain 670 AI sortase proteins are predicted to be involved in the secretion and anchoring of the LPXTG containing surface proteins. *S. pneumoniae* strain 670 AI may encode for at least one surface protein. Alternatively, *S. pneumoniae* strain 670 AI may encode for at least two surface exposed proteins and at least one sortase. Preferably, *S. pneumoniae* strain 670 AI encodes for at least three surface exposed proteins and at least two sortases.

The AI surface proteins may be covalently attached to the bacterial cell wall by membrane-associated transpeptidases, such as an AI sortase. The sortase may function to cleave the surface protein, preferably between the threonine and glycine residues of an LPXTG motif. The sortase may then assist in the formation of an amide link between the threonine carboxyl group and a cell wall precursor such as lipid II. The precursor can then be incorporated into the peptidoglycan via the transglycoslylation and transpeptidation reactions of bacterial wall synthesis. *See* Comfort et al., Infection & Immunity (2004) 72(5): 2710 – 2722.

In one embodiment, the invention includes a composition comprising oligomeric, pilus-like structures comprising a *S. pneumoniae* strain 670 AI surface protein such as orf3_670, orf4_670, or orf5_670. The oligomeric, pilus-like structure may comprise numerous units of AI surface protein. Preferably, the oligomeric, pilus-like structures comprise two or more AI surface proteins. Still more preferably, the oligomeric, pilus-like structure comprises a hyper-oligomeric pilus-like structure comprising at least two (*e.g.*, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 20, 25, 30, 35, 40, 45, 50, 60, 70, 80, 90, 100, 120, 140, 150, 200 or more) oligomeric subunits, wherein each subunit comprises an AI surface protein or a fragment thereof. The oligomeric subunits may be covalently associated via a conserved lysine within a pilin motif. The oligomeric subunits may be covalently associated via an LPXTG motif, preferably, via the threonine or serine amino acid residue, respectively.

AI surface proteins or fragments thereof to be incorporated into the oligomeric, pilus-like structures of the invention will preferably include a pilin motif.

The oligomeric, pilus like structures may be used alone or in the combinations of the invention. In one embodiment, the invention comprises a *S. pneumoniae* strain 670 AI protein in oligomeric form, preferably in a hyperoligomeric form. In one embodiment, the invention comprises a composition comprising one or more *S. pneumoniae* strain 670 AI proteins and one or more *S. pneumoniae* from TIGR4 AI proteins, wherein one or more of the *S. pneumoniae* AI proteins is in the form of an oligomer, preferably in a hyperoligomeric form.

In addition to the open reading frames encoding the *S. pneumoniae* strain 670 AI proteins, *S. pneumoniae* strain 670 AI may also include a transcriptional regulator.

S. pneumoniae strain 14 CSR 10 Adhesin Island

As discussed above, Applicants have identified adhesin islands within the genome of *S. pneumoniae* strain 14 CSR 10. The *S. pneumoniae* strain 14 CSR 10 Adhesin Island comprises a series of approximately seven open reading frames encoding for a collection of amino acid sequences comprising surface proteins and sortases. Specifically, the *S. pneumoniae* strain 14 CSR 10 AI proteins includes open reading frames encoding for two or more (*i.e.*, 2, 3, 4, 5, 6, or 7) of ORF2_14CSR, ORF3_14CSR, ORF4_14CSR, ORF5_14CSR, ORF6_14CSR, ORF7_14CSR, ORF8_14CSR.

A preferred immunogenic composition of the invention comprises a *S. pneumoniae* strain 14 CSR 10 AI surface protein which may be formulated or purified in an oligomeric (pilis) form. Another preferred immunogenic composition of the invention comprises a *S. pneumoniae* strain 14 CSR 10 AI surface protein which has been isolated in an oligomeric (pilis) form.

One or more of the *S. pneumoniae* strain 14 CSR 10 AI open reading frame polynucleotide sequences may be replaced by a polynucleotide sequence coding for a fragment of the replaced ORF. Alternatively, one or more of the *S. pneumoniae* strain 14 CSR 10 AI open reading frames may be replaced by a sequence having sequence homology to the replaced ORF.

One or more of the *S. pneumoniae* strain 14 CSR 10 AI surface protein sequences typically include an LPXTG motif (such as LPXTG (SEQ ID NO: 122)) or other sortase substrate motif.

The *S. pneumoniae* strain 14 CSR 10 AI surface proteins of the invention may affect the ability of the *S. pneumoniae* bacteria to adhere to and invade epithelial cells. AI surface proteins may also affect the ability of *S. pneumoniae* to translocate through an epithelial cell layer. Preferably, one or more *S. pneumoniae* strain 14 CSR 10 AI surface proteins are capable of binding to or otherwise associating with an epithelial cell surface. *S. pneumoniae* strain 14 CSR 10 AI surface proteins may also be able to bind to or associate with fibrinogen, fibronectin, or collagen.

The *S. pneumoniae* strain 14 CSR 10 AI sortase proteins are predicted to be involved in the secretion and anchoring of the LPXTG containing surface proteins. *S. pneumoniae* strain 14 CSR 10 AI may encode for at least one surface protein. Alternatively, *S. pneumoniae* strain 14 CSR 10 AI may encode for at least two surface exposed proteins and at least one sortase. Preferably, *S. pneumoniae* strain 14 CSR 10 AI encodes for at least three surface exposed proteins and at least two sortases.

The AI surface proteins may be covalently attached to the bacterial cell wall by membrane-associated transpeptidases, such as an AI sortase. The sortase may function to cleave the surface protein, preferably between the threonine and glycine residues of an LPXTG motif. The sortase may then assist in the formation of an amide link between the threonine carboxyl group and a cell wall precursor such as lipid II. The precursor can then be incorporated into the peptidoglycan via the transglycoslylation and transpeptidation reactions of bacterial wall synthesis. *See* Comfort et al., Infection & Immunity (2004) 72(5): 2710 – 2722.

In one embodiment, the invention includes a composition comprising oligomeric, pilus-like structures comprising a *S. pneumoniae* strain 14 CSR 10 AI surface protein such as orf3_CSR, orf4_CSR, or orf5_CSR. The oligomeric, pilus-like structure may comprise numerous units of AI surface protein. Preferably, the oligomeric, pilus-like structures comprise two or more AI surface proteins. Still more preferably, the oligomeric, pilus-like structure comprises a hyper-oligomeric pilus-like structure comprising at least two (*e.g.*, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 20, 25, 30, 35, 40, 45, 50, 60, 70, 80, 90, 100, 120, 140, 150, 200 or more) oligomeric subunits, wherein each subunit comprises an AI surface protein or a fragment thereof. The oligomeric subunits may be covalently associated via a conserved lysine within a pilin motif. The oligomeric subunits may be covalently associated via an LPXTG motif, preferably, via the threonine or serine amino acid residue, respectively.

AI surface proteins or fragments thereof to be incorporated into the oligomeric, pilus-like structures of the invention will preferably include a pilin motif.

The oligomeric, pilus like structures may be used alone or in the combinations of the invention. In one embodiment, the invention comprises a *S. pneumoniae* strain 14 CSR 10 AI protein in oligomeric form, preferably in a hyperoligomeric form. In one embodiment, the invention comprises a composition comprising one or more *S. pneumoniae* strain 14 CSR 10 AI proteins, and one or more AI proteins of any of *S. pneumoniae* from TIGR4, 670, 19A Hungary 6, 6B Finland 12, 6B Spain 2, 9V Spain 3, 19F Taiwan 14, 23F Taiwan 15, or 23F Poland 16, wherein one or more of the *S. pneumoniae* AI proteins is in the form of an oligomer, preferably in a hyperoligomeric form.

In addition to the open reading frames encoding the *S. pneumoniae* strain 14 CSR 10AI proteins, *S. pneumoniae* strain 14 CSR 10 AI may also include a transcriptional regulator.

S. pneumoniae strain 19A Hungary 6 Adhesin Island

As discussed above, Applicants have identified adhesin islands within the genome of *S. pneumoniae* strain 19A Hungary 6. The *S. pneumoniae* strain 19A Hungary 6 Adhesin Island comprises a series of approximately seven open reading frames encoding for a collection of amino acid sequences comprising surface proteins and sortases. Specifically, the *S. pneumoniae* strain 19A Hungary 6 AI proteins includes open reading frames encoding for two or more (*i.e.*, 2, 3, 4, 5, 6, or 7) of ORF2_19AH, ORF3_19AH, ORF4_19AH, ORF5_19AH, ORF6_19AH, ORF7_19AH, ORF8_19AH.

A preferred immunogenic composition of the invention comprises a *S. pneumoniae* strain 19A Hungary 6 AI surface protein which may be formulated or purified in an oligomeric (pilis) form. Another preferred immunogenic composition of the invention comprises a *S. pneumoniae* strain 19A Hungary 6 AI surface protein which has been isolated in an oligomeric (pilis) form.

One or more of the *S. pneumoniae* strain 19A Hungary 6 AI open reading frame polynucleotide sequences may be replaced by a polynucleotide sequence coding for a fragment of the replaced ORF. Alternatively, one or more of the *S. pneumoniae* strain 19A Hungary 6 AI open reading frames may be replaced by a sequence having sequence homology to the replaced ORF.

One or more of the *S. pneumoniae* strain 19A Hungary 6 AI surface protein sequences typically include an LPXTG motif (such as LPXTG (SEQ ID NO: 122)) or other sortase substrate motif.

The *S. pneumoniae* strain 19A Hungary 6 AI surface proteins of the invention may affect the ability of the *S. pneumoniae* bacteria to adhere to and invade epithelial cells. AI surface proteins may also affect the ability of *S. pneumoniae* to translocate through an epithelial cell layer. Preferably, one or more *S. pneumoniae* strain 19A Hungary 6 AI surface proteins are capable of binding to or otherwise associating with an epithelial cell surface. *S. pneumoniae* strain 19A Hungary 6 AI surface proteins may also be able to bind to or associate with fibrinogen, fibronectin, or collagen.

The *S. pneumoniae* strain 19A Hungary 6 AI sortase proteins are predicted to be involved in the secretion and anchoring of the LPXTG containing surface proteins. *S. pneumoniae* strain 19A Hungary 6 AI may encode for at least one surface protein. Alternatively, *S. pneumoniae* strain 19A Hungary 6 AI may encode for at least two surface exposed proteins and at least one sortase.

Preferably, *S. pneumoniae* strain 19A Hungary 6 AI encodes for at least three surface exposed proteins and at least two sortases.

The AI surface proteins may be covalently attached to the bacterial cell wall by membrane-associated transpeptidases, such as an AI sortase. The sortase may function to cleave the surface protein, preferably between the threonine and glycine residues of an LPXTG motif. The sortase may then assist in the formation of an amide link between the threonine carboxyl group and a cell wall precursor such as lipid II. The precursor can then be incorporated into the peptidoglycan via the transglycoslylation and transpeptidation reactions of bacterial wall synthesis. *See* Comfort et al., Infection & Immunity (2004) 72(5): 2710 – 2722.

In one embodiment, the invention includes a composition comprising oligomeric, pilus-like structures comprising a *S. pneumoniae* strain 19A Hungary 6 AI surface protein such as orf3_19AH, orf4_19AH, or orf5_19AH. The oligomeric, pilus-like structure may comprise numerous units of AI surface protein. Preferably, the oligomeric, pilus-like structures comprise two or more AI surface proteins. Still more preferably, the oligomeric, pilus-like structure comprises a hyper-oligomeric pilus-like structure comprising at least two (*e.g.*, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 20, 25, 30, 35, 40, 45, 50, 60, 70, 80, 90, 100, 120, 140, 150, 200 or more) oligomeric subunits, wherein each subunit comprises an AI surface protein or a fragment thereof. The oligomeric subunits may be covalently associated via a conserved lysine within a pilin motif. The oligomeric subunits may be covalently associated via an LPXTG motif, preferably, via the threonine or serine amino acid residue, respectively.

AI surface proteins or fragments thereof to be incorporated into the oligomeric, pilus-like structures of the invention will preferably include a pilin motif.

The oligomeric, pilus like structures may be used alone or in the combinations of the invention. In one embodiment, the invention comprises a *S. pneumoniae* strain 19A Hungary 6 AI protein in oligomeric form, preferably in a hyperoligomeric form. In one embodiment, the invention comprises a composition comprising one or more *S. pneumoniae* strain 19A Hungary 6 AI proteins and one or more AI proteins from one of any one of *S. pneumoniae* from TIGR4, 670, 14 CSR 10, 6B Finland 12, 6B Spain 2, 9V Spain 3, 19F Taiwan 14, 23F Taiwan 15, or 23F Poland 16 AI GR4 AI proteins, wherein one or more of the *S. pneumoniae* AI proteins is in the form of an oligomer, preferably in a hyperoligomeric form.

In addition to the open reading frames encoding the *S. pneumoniae* strain 19A Hungary 6 AI proteins, *S. pneumoniae* strain 19A Hungary 6 AI may also include a transcriptional regulator.

S. pneumoniae strain 19F Taiwan 14 Adhesin Island

As discussed above, Applicants have identified adhesin islands within the genome of *S. pneumoniae* strain 19F Taiwan 14. The *S. pneumoniae* strain 19F Taiwan 14 Adhesin Island comprises a series of approximately seven open reading frames encoding for a collection of amino acid sequences comprising surface proteins and sortases. Specifically, the *S. pneumoniae* strain 19F Taiwan 14 AI proteins includes open reading frames encoding for two or more (*i.e.*, 2, 3, 4, 5, 6, or 7) of ORF2_19FTW, ORF3_19FTW, ORF4_19FTW, ORF5_19FTW, ORF6_19FTW, ORF7_19FTW, ORF8_19FTW.

A preferred immunogenic composition of the invention comprises a *S. pneumoniae* strain 19F Taiwan 14 AI surface protein which may be formulated or purified in an oligomeric (pilis) form. Another preferred immunogenic composition of the invention comprises a *S. pneumoniae* strain 19F Taiwan 14 AI surface protein which has been isolated in an oligomeric (pilis) form.

One or more of the *S. pneumoniae* strain 19F Taiwan 14 AI open reading frame polynucleotide sequences may be replaced by a polynucleotide sequence coding for a fragment of the replaced ORF. Alternatively, one or more of the *S. pneumoniae* strain 19F Taiwan 14 AI open reading frames may be replaced by a sequence having sequence homology to the replaced ORF.

One or more of the *S. pneumoniae* strain 19F Taiwan 14 AI surface protein sequences typically include an LPXTG motif (such as LPXTG (SEQ ID NO: 122)) or other sortase substrate motif.

The *S. pneumoniae* strain 19F Taiwan 14 AI surface proteins of the invention may affect the ability of the *S. pneumoniae* bacteria to adhere to and invade epithelial cells. AI surface proteins may also affect the ability of *S. pneumoniae* to translocate through an epithelial cell layer. Preferably, one or more *S. pneumoniae* strain 19F Taiwan 14 AI surface proteins are capable of binding to or otherwise associating with an epithelial cell surface. *S. pneumoniae* strain 19F Taiwan 14 AI surface proteins may also be able to bind to or associate with fibrinogen, fibronectin, or collagen.

The *S. pneumoniae* strain 19F Taiwan 14 AI sortase proteins are predicted to be involved in the secretion and anchoring of the LPXTG containing surface proteins. *S. pneumoniae* strain 19F Taiwan 14 AI may encode for at least one surface protein. Alternatively, *S. pneumoniae* strain 19F Taiwan 14 AI may encode for at least two surface exposed proteins and at least one sortase. Preferably, *S. pneumoniae* strain 19F Taiwan 14 AI encodes for at least three surface exposed proteins and at least two sortases.

The AI surface proteins may be covalently attached to the bacterial cell wall by membrane-associated transpeptidases, such as an AI sortase. The sortase may function to cleave the surface protein, preferably between the threonine and glycine residues of an LPXTG motif. The sortase may then assist in the formation of an amide link between the threonine carboxyl group and a cell wall precursor such as lipid II. The precursor can then be incorporated into the peptidoglycan via the transglycoslylation and transpeptidation reactions of bacterial wall synthesis. *See* Comfort et al., Infection & Immunity (2004) 72(5): 2710 – 2722.

In one embodiment, the invention includes a composition comprising oligomeric, pilus-like structures comprising a *S. pneumoniae* strain 19F Taiwan 14 AI surface protein such as orf3_19FTW, orf4_19FTW, or orf5_19FTW. The oligomeric, pilus-like structure may comprise numerous units of AI surface protein. Preferably, the oligomeric, pilus-like structures comprise two or more AI surface proteins. Still more preferably, the oligomeric, pilus-like structure comprises a hyper-oligomeric pilus-like structure comprising at least two (*e.g.*, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 20, 25, 30, 35, 40, 45, 50, 60, 70, 80, 90, 100, 120, 140, 150, 200 or more) oligomeric subunits, wherein each subunit comprises an AI surface protein or a fragment thereof. The oligomeric subunits may be covalently associated via a conserved lysine within a pilin motif. The oligomeric subunits may be covalently associated via an LPXTG motif, preferably, via the threonine or serine amino acid residue, respectively.

AI surface proteins or fragments thereof to be incorporated into the oligomeric, pilus-like structures of the invention will preferably include a pilin motif.

The oligomeric, pilus like structures may be used alone or in the combinations of the invention. In one embodiment, the invention comprises a *S. pneumoniae* strain 19F Taiwan 14 AI protein in oligomeric form, preferably in a hyperoligomeric form. In one embodiment, the invention comprises a composition comprising one or more *S. pneumoniae* strain 19F Taiwan 14 AI proteins and one or more AI proteins of any one or more of *S. pneumoniae* from TIGR4, 670, 19A Hungary 6, 6B Finland 12, 6B Spain 2, 9V Spain 3, 14 CSR 10, 23F Taiwan 15, or 23F Poland 16, wherein one or more of the *S. pneumoniae* AI proteins is in the form of an oligomer, preferably in a hyperoligomeric form.

In addition to the open reading frames encoding the *S. pneumoniae* strain 19F Taiwan 14 AI proteins, *S. pneumoniae* strain 19F Taiwan 14 AI may also include a transcriptional regulator.

S. pneumoniae strain 23F Poland 16 Adhesin Island

As discussed above, Applicants have identified adhesin islands within the genome of *S. pneumoniae* strain 23F Poland 16. The *S. pneumoniae* strain 23F Poland 16 Adhesin Island comprises a series of approximately seven open reading frames encoding for a collection of amino acid sequences comprising surface proteins and sortases. Specifically, the *S. pneumoniae* strain 23F Poland 16 AI proteins includes open reading frames encoding for two or more (*i.e.*, 2, 3, 4, 5, 6, or 7) of ORF2_23FP, ORF3_23FP, ORF4_23FP, ORF5_23FP, ORF6_23FP, ORF7_23FP, and ORF8_23FP.

A preferred immunogenic composition of the invention comprises a *S. pneumoniae* strain 23F Poland 16 AI surface protein which may be formulated or purified in an oligomeric (pilis) form. Another preferred immunogenic composition of the invention comprises a *S. pneumoniae* strain 23F Poland 16 AI surface protein which has been isolated in an oligomeric (pilis) form.

One or more of the *S. pneumoniae* strain 23F Poland 16 AI open reading frame polynucleotide sequences may be replaced by a polynucleotide sequence coding for a fragment of the replaced ORF. Alternatively, one or more of the *S. pneumoniae* strain 23F Poland 16 AI open reading frames may be replaced by a sequence having sequence homology to the replaced ORF.

One or more of the *S. pneumoniae* strain 23F Poland 16 AI surface protein sequences typically include an LPXTG motif (such as LPXTG (SEQ ID NO: 122)) or other sortase substrate motif.

The *S. pneumoniae* strain 23F Poland 16 AI surface proteins of the invention may affect the ability of the *S. pneumoniae* bacteria to adhere to and invade epithelial cells. AI surface proteins may also affect the ability of *S. pneumoniae* to translocate through an epithelial cell layer. Preferably, one or more *S. pneumoniae* strain 23F Poland 16 AI surface proteins are capable of binding to or otherwise associating with an epithelial cell surface. *S. pneumoniae* strain 23F Poland 16 AI surface proteins may also be able to bind to or associate with fibrinogen, fibronectin, or collagen.

The *S. pneumoniae* strain 23F Poland 16 AI sortase proteins are predicted to be involved in the secretion and anchoring of the LPXTG containing surface proteins. *S. pneumoniae* strain 23F Poland 16 AI may encode for at least one surface protein. Alternatively, *S. pneumoniae* strain 23F Poland 16 AI may encode for at least two surface exposed proteins and at least one sortase. Preferably, *S. pneumoniae* strain 23F Poland 16 AI encodes for at least three surface exposed proteins and at least two sortases.

The AI surface proteins may be covalently attached to the bacterial cell wall by membrane-associated transpeptidases, such as an AI sortase. The sortase may function to cleave the surface protein, preferably between the threonine and glycine residues of an LPXTG motif. The sortase may then assist in the formation of an amide link between the threonine carboxyl group and a cell wall precursor such as lipid II. The precursor can then be incorporated into the peptidoglycan via the transglycoslylation and transpeptidation reactions of bacterial wall synthesis. *See* Comfort et al., Infection & Immunity (2004) 72(5): 2710 – 2722.

In one embodiment, the invention includes a composition comprising oligomeric, pilus-like structures comprising a *S. pneumoniae* strain 23F Poland 16 AI surface protein such as orf3_23FP, orf4_23FP, or orf5_23FP. The oligomeric, pilus-like structure may comprise numerous units of AI surface protein. Preferably, the oligomeric, pilus-like structures comprise two or more AI surface proteins. Still more preferably, the oligomeric, pilus-like structure comprises a hyper-oligomeric pilus-like structure comprising at least two (*e.g.*, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 20, 25, 30, 35, 40, 45, 50, 60, 70, 80, 90, 100, 120, 140, 150, 200 or more) oligomeric subunits, wherein each subunit comprises an AI surface protein or a fragment thereof. The oligomeric subunits may be covalently associated via a conserved lysine within a pilin motif. The oligomeric subunits may be covalently associated via an LPXTG motif, preferably, via the threonine or serine amino acid residue, respectively.

AI surface proteins or fragments thereof to be incorporated into the oligomeric, pilus-like structures of the invention will preferably include a pilin motif.

The oligomeric, pilus like structures may be used alone or in the combinations of the invention. In one embodiment, the invention comprises a *S. pneumoniae* strain 23F Poland 16 AI protein in oligomeric form, preferably in a hyperoligomeric form. In one embodiment, the invention comprises a composition comprising one or more *S. pneumoniae* strain 23F Poland 16 AI proteins and one or more AI proteins from any one or more *S. pneumoniae* strains of TIGR4, 670, 19A Hungary 6, 6B Finland 12, 6B Spain 2, 9V Spain 3, 19F Taiwan 14, 23F Taiwan 15, or 14 CSR 10, wherein one or more of the *S. pneumoniae* AI proteins is in the form of an oligomer, preferably in a hyperoligomeric form.

In addition to the open reading frames encoding the *S. pneumoniae* strain 23F Poland 16 AI proteins, *S. pneumoniae* strain 23F Poland 16 AI may also include a transcriptional regulator.

<u>*S. pneumoniae* strain 23F Taiwan 15 Adhesin Island</u>

As discussed above, Applicants have identified adhesin islands within the genome of *S. pneumoniae* strain 23F Taiwan 15. The *S. pneumoniae* strain 23F Taiwan 15 Adhesin Island comprises a series of approximately seven open reading frames encoding for a collection of amino acid sequences comprising surface proteins and sortases. Specifically, the *S. pneumoniae* strain 23F Taiwan 15 AI proteins includes open reading frames encoding for two or more (*i.e.*, 2, 3, 4, 5, 6, or 7) of ORF2_23FTW, ORF3_23FTW, ORF4_23FTW, ORF5_23FTW, ORF6_23FTW, ORF7_23FTW, ORF8_23FTW.

A preferred immunogenic composition of the invention comprises a *S. pneumoniae* strain 23F Taiwan 15 AI surface protein which may be formulated or purified in an oligomeric (pilis) form. Another preferred immunogenic composition of the invention comprises a *S. pneumoniae* strain 23F Taiwan 15 AI surface protein which has been isolated in an oligomeric (pilis) form.

One or more of the *S. pneumoniae* strain 23F Taiwan 15 AI open reading frame polynucleotide sequences may be replaced by a polynucleotide sequence coding for a fragment of the replaced ORF. Alternatively, one or more of the *S. pneumoniae* strain 23F Taiwan 15 AI open reading frames may be replaced by a sequence having sequence homology to the replaced ORF.

One or more of the *S. pneumoniae* strain 23F Taiwan 15 AI surface protein sequences typically include an LPXTG motif (such as LPXTG (SEQ ID NO: 122)) or other sortase substrate motif.

The *S. pneumoniae* strain 23F Taiwan 15 AI surface proteins of the invention may affect the ability of the *S. pneumoniae* bacteria to adhere to and invade epithelial cells. AI surface proteins may also affect the ability of *S. pneumoniae* to translocate through an epithelial cell layer. Preferably, one or more *S. pneumoniae* strain 23F Taiwan 15 AI surface proteins are capable of binding to or otherwise associating with an epithelial cell surface. *S. pneumoniae* strain 23F Taiwan 15 AI surface proteins may also be able to bind to or associate with fibrinogen, fibronectin, or collagen.

The *S. pneumoniae* strain 23F Taiwan 15 AI sortase proteins are predicted to be involved in the secretion and anchoring of the LPXTG containing surface proteins. *S. pneumoniae* strain 23F Taiwan 15 AI may encode for at least one surface protein. Alternatively, *S. pneumoniae* strain 23F Taiwan 15 AI may encode for at least two surface exposed proteins and at least one sortase. Preferably, *S. pneumoniae* strain 23F Taiwan 15 AI encodes for at least three surface exposed proteins and at least two sortases.

The AI surface proteins may be covalently attached to the bacterial cell wall by membrane-associated transpeptidases, such as an AI sortase. The sortase may function to cleave the surface protein, preferably between the threonine and glycine residues of an LPXTG motif. The sortase may then assist in the formation of an amide link between the threonine carboxyl group and a cell wall precursor such as lipid II. The precursor can then be incorporated into the peptidoglycan via the transglycosylation and transpeptidation reactions of bacterial wall synthesis. *See* Comfort et al., Infection & Immunity (2004) 72(5): 2710 – 2722.

In one embodiment,.the invention includes a composition comprising oligomeric, pilus-like structures comprising a *S. pneumoniae* strain 23F Taiwan 15 AI surface protein such as orf3_23FTW, orf4_23FTW, or orf5_23FTW. The oligomeric, pilus-like structure may comprise numerous units of AI surface protein. Preferably, the oligomeric, pilus-like structures comprise two or more AI surface proteins. Still more preferably, the oligomeric, pilus-like structure comprises a hyper-oligomeric pilus-like structure comprising at least two (*e.g.*, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 20, 25, 30, 35, 40, 45, 50, 60, 70, 80, 90, 100, 120, 140, 150, 200 or more) oligomeric subunits, wherein each subunit comprises an AI surface protein or a fragment thereof. The oligomeric subunits may be covalently associated via a conserved lysine within a pilin motif. The oligomeric subunits may be covalently associated via an LPXTG motif, preferably, via the threonine or serine amino acid residue, respectively.

AI surface proteins or fragments thereof to be incorporated into the oligomeric, pilus-like structures of the invention will preferably include a pilin motif.

The oligomeric, pilus like structures may be used alone or in the combinations of the invention. In one embodiment, the invention comprises a *S. pneumoniae* strain 23F Taiwan 15 AI protein in oligomeric form, preferably in a hyperoligomeric form. In one embodiment, the invention comprises a composition comprising one or more *S. pneumoniae* strain 23F Taiwan 15 AI proteins and one or more AI proteins from any one or more of *S. pneumoniae* from TIGR4, 670, 19A Hungary 6, 6B Finland 12, 6B Spain 2, 9V Spain 3, 19F Taiwan 14, 14 CSR 10, or 23F Poland 16 AI, wherein one or more of the *S. pneumoniae* AI proteins is in the form of an oligomer, preferably in a hyperoligomeric form.

In addition to the open reading frames encoding the *S. pneumoniae* strain 23F Taiwan 15 AI proteins, *S. pneumoniae* strain 23F Taiwan 15 AI may also include a transcriptional regulator.

S. pneumoniae strain 6B Finland 12 Adhesin Island

As discussed above, Applicants have identified adhesin islands within the genome of *S. pneumoniae* strain 6B Finland 12. The *S. pneumoniae* strain 6B Finland 12 Adhesin Island comprises a series of approximately seven open reading frames encoding for a collection of amino acid sequences comprising surface proteins and sortases. Specifically, the *S. pneumoniae* strain 6B Finland 12 AI proteins includes open reading frames encoding for two or more (*i.e.*, 2, 3, 4, 5, 6, or 7) of ORF2_6BF, ORF3_6BF, ORF4_6BF, ORF5_6BF, ORF6_6BF, ORF7_6BF, ORF8_6BF.

A preferred immunogenic composition of the invention comprises a *S. pneumoniae* strain 6B Finland 12 AI surface protein which may be formulated or purified in an oligomeric (pilis) form. Another preferred immunogenic composition of the invention comprises a *S. pneumoniae* strain 6B Finland 12 AI surface protein which has been isolated in an oligomeric (pilis) form.

One or more of the *S. pneumoniae* strain 6B Finland 12 AI open reading frame polynucleotide sequences may be replaced by a polynucleotide sequence coding for a fragment of the replaced ORF.

Alternatively, one or more of the *S. pneumoniae* strain 6B Finland 12 AI open reading frames may be replaced by a sequence having sequence homology to the replaced ORF.

One or more of the *S. pneumoniae* strain 6B Finland 12 AI surface protein sequences typically include an LPXTG motif (such as LPXTG (SEQ ID NO: 122)) or other sortase substrate motif.

The *S. pneumoniae* strain 6B Finland 12 AI surface proteins of the invention may affect the ability of the *S. pneumoniae* bacteria to adhere to and invade epithelial cells. AI surface proteins may also affect the ability of *S. pneumoniae* to translocate through an epithelial cell layer. Preferably, one or more *S. pneumoniae* strain 6B Finland 12 AI surface proteins are capable of binding to or otherwise associating with an epithelial cell surface. *S. pneumoniae* strain 6B Finland 12 AI surface proteins may also be able to bind to or associate with fibrinogen, fibronectin, or collagen.

The *S. pneumoniae* strain 6B Finland 12 AI sortase proteins are predicted to be involved in the secretion and anchoring of the LPXTG containing surface proteins. *S. pneumoniae* strain 6B Finland 12 AI may encode for at least one surface protein. Alternatively, *S. pneumoniae* strain 6B Finland 12 AI may encode for at least two surface exposed proteins and at least one sortase. Preferably, *S. pneumoniae* strain 6B Finland 12 AI encodes for at least three surface exposed proteins and at least two sortases.

The AI surface proteins may be covalently attached to the bacterial cell wall by membrane-associated transpeptidases, such as an AI sortase. The sortase may function to cleave the surface protein, preferably between the threonine and glycine residues of an LPXTG motif. The sortase may then assist in the formation of an amide link between the threonine carboxyl group and a cell wall precursor such as lipid II. The precursor can then be incorporated into the peptidoglycan via the transglycoslylation and transpeptidation reactions of bacterial wall synthesis. *See* Comfort et al., Infection & Immunity (2004) 72(5): 2710 – 2722.

In one embodiment, the invention includes a composition comprising oligomeric, pilus-like structures comprising a *S. pneumoniae* strain 6B Finland 12 AI surface protein such as orf3_6BF, orf4_6BF, or orf5_6BF. The oligomeric, pilus-like structure may comprise numerous units of AI surface protein. Preferably, the oligomeric, pilus-like structures comprise two or more AI surface proteins. Still more preferably, the oligomeric, pilus-like structure comprises a hyper-oligomeric pilus-like structure comprising at least two (*e.g.*, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 20, 25, 30, 35, 40, 45, 50, 60, 70, 80, 90, 100, 120, 140, 150, 200 or more) oligomeric subunits, wherein each subunit comprises an AI surface protein or a fragment thereof. The oligomeric subunits may be covalently associated via a conserved lysine within a pilin motif. The oligomeric subunits may be covalently associated via an LPXTG motif, preferably, via the threonine or serine amino acid residue, respectively.

AI surface proteins or fragments thereof to be incorporated into the oligomeric, pilus-like structures of the invention will preferably include a pilin motif.

The oligomeric, pilus-like structures may be used alone or in the combinations of the invention. In one embodiment, the invention comprises a *S. pneumoniae* strain 6B Finland 12 AI protein in oligomeric form, preferably in a hyperoligomeric form. In one embodiment, the invention comprises a composition comprising one or more *S. pneumoniae* strain 6B Finland 12 AI proteins and one or more AI proteins of any one or more of *S. pneumoniae* from TIGR4, 670, 19A Hungary 6, 6B Finland 12, 6B Spain 2, 9V Spain 3, 19F Taiwan 14, 23F Taiwan 15, or 23F Poland 16 AI, wherein one or more of the *S. pneumoniae* AI proteins is in the form of an oligomer, preferably in a hyperoligomeric form.

In addition to the open reading frames encoding the *S. pneumoniae* strain 6B Finland 12 AI proteins, *S. pneumoniae* strain 6B Finland 12 AI may also include a transcriptional regulator.

<u>*S. pneumoniae* strain 6B Spain 2 Adhesin Island</u>

As discussed above, Applicants have identified adhesin islands within the genome of *S. pneumoniae* strain 6B Spain 2. The *S. pneumoniae* strain 6B Spain 2 Adhesin Island comprises a series of approximately seven open reading frames encoding for a collection of amino acid sequences comprising surface proteins and sortases. Specifically, the *S. pneumoniae* strain 6B Spain 2 AI proteins includes open reading frames encoding for two or more (*i.e.*, 2, 3, 4, 5, 6, or 7) of ORF2_6BSP, ORF3_6BSP, ORF4_6BSP, ORF5_6BSP, ORF6_6BSP, ORF7_6BSP, and ORF8_6BSP.

A preferred immunogenic composition of the invention comprises a *S. pneumoniae* strain 6B Spain 2 AI surface protein which may be formulated or purified in an oligomeric (pilis) form. Another preferred immunogenic composition of the invention comprises a *S. pneumoniae* strain 6B Spain 2 AI surface protein which has been isolated in an oligomeric (pilis) form.

One or more of the *S. pneumoniae* strain 6B Spain 2 AI open reading frame polynucleotide sequences may be replaced by a polynucleotide sequence coding for a fragment of the replaced ORF. Alternatively, one or more of the *S. pneumoniae* strain 6B Spain 2 AI open reading frames may be replaced by a sequence having sequence homology to the replaced ORF.

One or more of the *S. pneumoniae* strain 6B Spain 2 AI surface protein sequences typically include an LPXTG motif (such as LPXTG (SEQ ID NO: 122)) or other sortase substrate motif.

The *S. pneumoniae* strain 6B Spain 2 AI surface proteins of the invention may affect the ability of the *S. pneumoniae* bacteria to adhere to and invade epithelial cells. AI surface proteins may also affect the ability of *S. pneumoniae* to translocate through an epithelial cell layer. Preferably, one or more *S. pneumoniae* strain 6B Spain 2 AI surface proteins are capable of binding to or otherwise associating with an epithelial cell surface. *S. pneumoniae* strain 6B Spain 2 AI surface proteins may also be able to bind to or associate with fibrinogen, fibronectin, or collagen.

The *S. pneumoniae* strain 6B Spain 2 AI sortase proteins are predicted to be involved in the secretion and anchoring of the LPXTG containing surface proteins. *S. pneumoniae* strain 6B Spain 2 AI may encode for at least one surface protein. Alternatively, *S. pneumoniae* strain 6B Spain 2 AI may encode for at least two surface exposed proteins and at least one sortase. Preferably, *S.*

pneumoniae strain 6B Spain 2 AI encodes for at least three surface exposed proteins and at least two sortases.

The AI surface proteins may be covalently attached to the bacterial cell wall by membrane-associated transpeptidases, such as an AI sortase. The sortase may function to cleave the surface protein, preferably between the threonine and glycine residues of an LPXTG motif. The sortase may then assist in the formation of an amide link between the threonine carboxyl group and a cell wall precursor such as lipid II. The precursor can then be incorporated into the peptidoglycan via the transglycosylation and transpeptidation reactions of bacterial wall synthesis. *See* Comfort et al., Infection & Immunity (2004) 72(5): 2710 – 2722.

In one embodiment, the invention includes a composition comprising oligomeric, pilus-like structures comprising a *S. pneumoniae* strain 6B Spain 2 AI surface protein such as orf3_6BSP, orf4_6BSP, or orf5_6BSP. The oligomeric, pilus-like structure may comprise numerous units of AI surface protein. Preferably, the oligomeric, pilus-like structures comprise two or more AI surface proteins. Still more preferably, the oligomeric, pilus-like structure comprises a hyper-oligomeric pilus-like structure comprising at least two (*e.g.*, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 20, 25, 30, 35, 40, 45, 50, 60, 70, 80, 90, 100, 120, 140, 150, 200 or more) oligomeric subunits, wherein each subunit comprises an AI surface protein or a fragment thereof. The oligomeric subunits may be covalently associated via a conserved lysine within a pilin motif. The oligomeric subunits may be covalently associated via an LPXTG motif, preferably, via the threonine or serine amino acid residue, respectively.

AI surface proteins or fragments thereof to be incorporated into the oligomeric, pilus-like structures of the invention will preferably include a pilin motif.

The oligomeric, pilus like structures may be used alone or in the combinations of the invention. In one embodiment, the invention comprises a *S. pneumoniae* strain 6B Spain 2 AI protein in oligomeric form, preferably in a hyperoligomeric form. In one embodiment, the invention comprises a composition comprising one or more *S. pneumoniae* strain 6B Spain 2 AI proteins and one or more AI proteins of any one or more of *S. pneumoniae* from TIGR4, 670, 19A Hungary 6, 6B Finland 12, 14 CSR 10, 9V Spain 3, 19F Taiwan 14, 23F Taiwan 15, or 23F Poland 16 AI, wherein one or more of the *S. pneumoniae* AI proteins is in the form of an oligomer, preferably in a hyperoligomeric form.

In addition to the open reading frames encoding the *S. pneumoniae* strain 6B Spain 2 AI proteins, *S. pneumoniae* strain 6B Spain 2 AI may also include a transcriptional regulator.

<u>*S. pneumoniae* strain 9V Spain 3 Adhesin Island</u>

As discussed above, Applicants have identified adhesin islands within the genome of *S. pneumoniae* strain 9V Spain 3. The *S. pneumoniae* strain 9V Spain 3 Adhesin Island comprises a series of approximately seven open reading frames encoding for a collection of amino acid sequences comprising surface proteins and sortases. Specifically, the *S. pneumoniae* strain 9V Spain 3 AI proteins includes open reading frames encoding for two or more (*i.e.*, 2, 3, 4, 5, 6, or 7) of

ORF2_9VSP, ORF3_9VSP, ORF4_9VSP, ORF5_9VSP, ORF6_9VSP, ORF7_9VSP, and ORF8_9VSP.

A preferred immunogenic composition of the invention comprises a *S. pneumoniae* strain 9V Spain 3 AI surface protein which may be formulated or purified in an oligomeric (pilis) form. Another preferred immunogenic composition of the invention comprises a *S. pneumoniae* strain 9V Spain 3 AI surface protein which has been isolated in an oligomeric (pilis) form.

One or more of the *S. pneumoniae* strain 9V Spain 3 AI open reading frame polynucleotide sequences may be replaced by a polynucleotide sequence coding for a fragment of the replaced ORF. Alternatively, one or more of the *S. pneumoniae* strain 9V Spain 3 AI open reading frames may be replaced by a sequence having sequence homology to the replaced ORF.

One or more of the *S. pneumoniae* strain 9V Spain 3 AI surface protein sequences typically include an LPXTG motif (such as LPXTG (SEQ ID NO: 122)) or other sortase substrate motif.

The *S. pneumoniae* strain 9V Spain 3 AI surface proteins of the invention may affect the ability of the *S. pneumoniae* bacteria to adhere to and invade epithelial cells. AI surface proteins may also affect the ability of *S. pneumoniae* to translocate through an epithelial cell layer. Preferably, one or more *S. pneumoniae* strain 9V Spain 3 AI surface proteins are capable of binding to or otherwise associating with an epithelial cell surface. *S. pneumoniae* strain 9V Spain 3 AI surface proteins may also be able to bind to or associate with fibrinogen, fibronectin, or collagen.

The *S. pneumoniae* strain 9V Spain 3 AI sortase proteins are predicted to be involved in the secretion and anchoring of the LPXTG containing surface proteins. *S. pneumoniae* strain 9V Spain 3 AI may encode for at least one surface protein. Alternatively, *S. pneumoniae* strain 9V Spain 3 AI may encode for at least two surface exposed proteins and at least one sortase. Preferably, *S. pneumoniae* strain 9V Spain 3 AI encodes for at least three surface exposed proteins and at least two sortases.

The AI surface proteins may be covalently attached to the bacterial cell wall by membrane-associated transpeptidases, such as an AI sortase. The sortase may function to cleave the surface protein, preferably between the threonine and glycine residues of an LPXTG motif. The sortase may then assist in the formation of an amide link between the threonine carboxyl group and a cell wall precursor such as lipid II. The precursor can then be incorporated into the peptidoglycan via the transglycoslylation and transpeptidation reactions of bacterial wall synthesis. *See* Comfort et al., Infection & Immunity (2004) 72(5): 2710 – 2722.

In one embodiment, the invention includes a composition comprising oligomeric, pilus-like structures comprising a *S. pneumoniae* strain 9V Spain 3 AI surface protein such as orf3_9VSP, orf4_9VSP, or orf5_9VSP. The oligomeric, pilus-like structure may comprise numerous units of AI surface protein. Preferably, the oligomeric, pilus-like structures comprise two or more AI surface proteins. Still more preferably, the oligomeric, pilus-like structure comprises a hyper-oligomeric pilus-like structure comprising at least two (*e.g.*, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 20, 25, 30, 35, 40, 45, 50, 60, 70, 80, 90, 100, 120, 140, 150, 200 or more) oligomeric subunits, wherein each subunit comprises an AI surface protein or a fragment thereof. The oligomeric subunits may be covalently associated via a conserved lysine within a pilin motif. The oligomeric subunits may be covalently associated via an LPXTG motif, preferably, via the threonine or serine amino acid residue, respectively.

AI surface proteins or fragments thereof to be incorporated into the oligomeric, pilus-like structures of the invention will preferably include a pilin motif.

The oligomeric, pilus like structures may be used alone or in the combinations of the invention. In one embodiment, the invention comprises a *S. pneumoniae* strain 9V Spain 3 AI protein in oligomeric form, preferably in a hyperoligomeric form. In one embodiment, the invention comprises a composition comprising one or more *S. pneumoniae* strain 9V Spain 3 AI proteins and one or more AI proteins from any one or more of *S. pneumoniae* from TIGR4, 670, 19A Hungary 6, 6B Finland 12, 6B Spain 2, 14 CSR 10, 19F Taiwan 14, 23F Taiwan 15, or 23F Poland 16 AI, wherein one or more of the *S. pneumoniae* AI proteins is in the form of an oligomer, preferably in a hyperoligomeric form.

In addition to the open reading frames encoding the *S. pneumoniae* strain 9V Spain 3 AI proteins, *S. pneumoniae* strain 9V Spain 3 AI may also include a transcriptional regulator.

The *S. pneumoniae* oligomeric, pilus-like structures may be isolated or purified from bacterial cultures in which the bacteria express an *S. pneumoniae* AI surface protein. The invention therefore includes a method for manufacturing an oligomeric AI surface antigen comprising culturing a *S. pneumoniae* bacterium that expresses the oligomeric AI protein and isolating the expressed oligomeric AI protein from the *S. pneumoniae* bacteria. The AI protein may be collected from secretions into the supernatant or it may be purified from the bacterial surface. The method may further comprise purification of the expressed AI protein. Preferably, the AI protein is in a hyperoligomeric form.

The oligomeric, pilus-like structures may be isolated or purified from bacterial cultures overexpressing an AI surface protein. The invention therefore includes a method for manufacturing an *S. pneumoniae* oligomeric Adhesin Island surface antigen comprising culturing a *S. pneumoniae* bacterium adapted for increased AI protein expression and isolation of the expressed oligomeric Adhesin Island protein from the *S. pneumoniae* bacteria. The AI protein may be collected from secretions into the supernatant or it may be purified from the bacterial surface. The method may further comprise purification of the expressed Adhesin Island protein. Preferably, the Adhesin Island protein is in a hyperoligomeric form.

The *S. pneumoniae* bacteria are preferably adapted to increase AI protein expression by at least two (*e.g.*, 2, 3, 4, 5, 8, 10, 15, 20, 25, 30, 35, 40, 45, 50, 60, 70, 80, 90, 100, 125, 150 or 200) times wild type expression levels.

S. pneumoniae bacteria may be adapted to increase AI protein expression by any means known in the art, including methods of increasing gene dosage and methods of gene upregulation. Such means include, for example, transformation of the *S. pneumoniae* bacteria with a plasmid encoding the AI protein. The plasmid may include a strong promoter or it may include multiple copies of the sequence encoding the AI protein. Optionally, the sequence encoding the AI protein within the *S. pneumoniae* bacterial genome may be deleted. Alternatively, or in addition, the promoter regulating the *S. pneumoniae* Adhesin Island may be modified to increase expression.

The invention further includes *S. pneumoniae* bacteria which have been adapted to produce increased levels of AI surface protein. In particular, the invention includes *S. pneumoniae* bacteria which have been adapted to produce oligomeric or hyperoligomeric AI surface protein. In one embodiment, the *S. pneumoniae* of the invention are inactivated or attenuated to permit *in vivo* delivery of the whole bacteria, with the AI surface protein exposed on its surface.

The invention further includes *S. pneumoniae* bacteria which have been adapted to have increased levels of expressed AI protein incorporated in pili on their surface. The *S. pneumoniae* bacteria may be adapted to have increased exposure of oligomeric or hyperoligomeric AI proteins on its surface by increasing expression levels of a signal peptidase polypeptide. Increased levels of a local signal peptidase expression in Gram positive bacteria (such us LepA in GAS) are expected to result in increased exposure of pili proteins on the surface of Gram positive bacteria. Increased expression of a leader peptidase in *S. pneumoniae* may be achieved by any means known in the art, such as increasing gene dosage and methods of gene upregulation. The *S. pneumoniae* bacteria adapted to have increased levels of leader peptidase may additionally be adapted to express increased levels of at least one pili protein.

Alternatively, the AI proteins of the invention may be expressed on the surface of a non-pathogenic Gram positive bacteria, such as *Streptococus gordonii* (See, e.g., Byrd et al., "Biological consequences of antigen and cytokine co-expression by recombinant *Streptococcus gordonii* vaccine vectors", Vaccine (2002) 20:2197-2205) or *Lactococcus lactis* (See, e.g., Mannam et al., "Mucosal Vaccine Made from Live, Recombinant Lactococcus lactis Protects Mice against Pharangeal Infection with *Streptococcus pyogenes*" Infection and Immunity (2004) 72(6):3444-3450). As used herein, non-pathogenic Gram positive bacteria refer to Gram positive bacteria which are compatible with a human host subject and are not associated with human pathogenisis. Preferably, the non-pathogenic bacteria are modified to express the AI surface protein in oligomeric, or hyper-oligomeric form. Sequences encoding for an AI surface protein and, optionally, an AI sortase, may be integrated into the non-pathogenic Gram positive bacterial genome or inserted into a plasmid. The non-pathogenic Gram positive bacteria may be inactivated or attenuated to facilitate *in vivo* delivery of the whole bacteria, with the AI surface protein exposed on its surface. Alternatively, the AI surface protein may be isolated or purified from a bacterial culture of the non-pathogenic Gram positive bacteria. For example, the AI surface protein may be isolated from cell extracts or culture supernatants. Alternatively, the AI surface protein may be isolated or purified from the surface of the non-pathogenic Gram positive bacteria.

The non-pathogenic Gram positive bacteria may be used to express any of the *S. pneumoniae* Adhesin Island proteins described herein. The non-pathogenic Gram positive bacteria are transformed to express an Adhesin Island surface protein. Preferably, the non-pathogenic Gram positive bacteria also express at least one Adhesin Island sortase. The AI transformed non-pathogenic Gram positive bacteria of the invention may be used to prevent or treat infection with pathogenic *S. pneumoniae.*

Figures 190 A and B, and 193-195 provide examples of three methods successfully practiced by applicants to purify pili from *S. pneumoniae* TIGR4.

Immunogenic Compositions

The Gram positive bacteria AI proteins described herein are useful in immunogenic compositions for the prevention or treatment of Gram positive bacterial infection. For example, the GBS AI surface proteins described herein are useful in immunogenic compositions for the prevention or treatment of GBS infection. As another example, the GAS AI surface proteins described herein may be useful in immunogenic compositions for the prevention or treatment of GAS infection. As another example, the *S. pneumoniae* AI surface proteins may be useful in immunogenic cojmpositions for the prevention or treatment of *S. pneumoniae* infection.

Gram positive bacteria AI surface proteins that can provide protection across more than one serotype or strain isolate may be used to increase immunogenic effectiveness. For example, a particular GBS AI surface protein having an amino acid sequence that is at least 50% (*i.e.*, at least 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100%) homologous to the particular GBS AI surface protein of at least 2 (*i.e.*, at least 3, 4, 5, 6, 7, 8, 9, 10, or more) other GBS serotypes or strain isolates may be used to increase the effectiveness of such compositions.

As another example, fragments of Gram positive bacteria AI surface proteins that can provide protection across more than one serotype or strain isolate may be used to increase immunogenic effectiveness. Such a fragment may be identified within a consensus sequence of a full length amino acid sequence of a Gram positive bacteria AI surface protein. Such a fragment can be identified in the consensus sequence by its high degree of homology or identity across multiple (*i.e*, at least 3, 4, 5, 6, 7, 8, 9, 10, or more) Gram positive bacteria serotypes or strain isolates. Preferably, a high degree of homology is a degree of homology of at least 90% (*i.e.*, at least 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or 100%) across Gram positive bacteria serotypes or strain isolates. Preferably, a high degree of identity is a degree of identity of at least 90% (*i.e.*, at least 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or 100%) across Gram positive bacteria serotypes or strain isolates. In one embodiment of the invention, such a fragment of a Gram positive bacteria AI surface protein may be used in the immunogenic compositions.

In addition, the AI surface protein oligomeric pilus structures may be formulated or purified for use in immunization. Isolated AI surface protein oligomeric pilus structures may also be used for immunization.

The invention includes an immunogenic composition comprising a first Gram positive bacteria AI protein and a second Gram positive bacterial AI protein. One or more of the AI proteins may be a surface protein. Such surface proteins may contain an LPXTG motif or other sortase substrate motif.

The first and second AI proteins may be from the same or different genus or species of Gram positive bacteria. If within the same species, the first and second AI proteins may be from the same or different AI subtypes. If two AIs are of the same subtype, the AIs have the same numerical designation. For example, all AIs designated as AI-1 are of the same AI subtype. If two AIs are of a different subtype, the AIs have different numerical designations. For example, AI-1 is of a different AI subtype from AI-2, AI-3, AI-4, etc. Likewise, AI-2 is of a different AI subtype from AI-1, AI-3, and AI-4, etc.

For example, the invention includes an immunogenic composition comprising one or more GBS AI-1 proteins and one or more GBS AI-2 proteins. One or more of the AI proteins may be a surface protein. Such surface proteins may contain an LPXTG motif (such as LPXTG (SEQ ID NO: 122)) and may bind fibrinogen, fibronectin, or collagen. One or more of the AI proteins may be a sortase. The GBS AI-1 proteins may be selected from the group consisting of GBS 80, GBS 104, GBS 52, SAG0647 and SAG0648. Preferably, the GBS AI-1 proteins include GBS 80 or GBS 104.

The GBS AI-2 proteins may be selected from the group consisting of GBS 67, GBS 59, GBS 150, SAG1405, SAG1406, 01520, 01521, 01522, 01523, 01523, 01524 and 01525. In one embodiment, the GBS AI-2 proteins are selected from the group consisting of GBS 67, GBS 59, GBS 150, SAG1405, and SAG1406. In another embodiment, the GBS AI-2 proteins may be selected from the group consisting of 01520, 01521, 01522, 01523, 01523, 01524 and 01525. Preferably, the GBS AI-2 protein includes GBS 59 or GBS 67.

As another example, the invention includes an immunogenic composition comprising one or more of any combination of GAS AI-1, GAS AI-2, GAS AI-3, or GAS AI-4 proteins. One or more of the GAS AI proteins may be a sortase. The GAS AI-1 proteins may be selected from the group consisting of M6_Spy0156, M6_Spy0157, M6_Spy0158, M6_Spy0159, M6_Spy0160, M6_Spy0161, CDC SS 410_fimbrial, ISS3650_fimbrial, and DSM2071_fimbrial. Preferably, the GAS AI-1 proteins are selected from the group consisting of M6_Spy0157, M6_Spy0159, M6_Spy0160, CDC SS 410_fimbrial, ISS3650_fimbrial, and DSM2071_fimbrial.

The GAS AI-2 proteins may be selected from the group consisting of Spy0124, GAS15, Spy0127, GAS16, GAS17, GAS18, Spy0131, Spy0133, and GAS20. Preferably, the GAS AI-2 proteins are selected from the group consisting of GAS15, GAS16, and GAS18.

The GAS AI-3 proteins may be selected from the group consisting of SpyM3_0097, SpyM3_0098, SpyM3_0099, SpyM3_0100, SpyM3_0101, SpyM3_0102, SpyM3_0103, SpyM3_0104, SPs0099, SPs0100, SPs0101, SPs0102, SPs0103, SPs0104, SPs0105, SPs0106, orf77, orf78, orf79, orf80, orf81, orf82, orf83, orf84, spyM18_0125, spyM18_0126, spyM18_0127, spyM18_0128, spyM18_0129, spyM18_0130, spyM18_0131, spyM18_0132, SpyoM01000156, SpyoM01000155, SpyoM01000154, SpyoM01000153, SpyoM01000152, SpyoM01000151, SpyoM01000150, SpyoM01000149, ISS3040_fimbrial, ISS3776_fimbrial, and ISS4959_fimbrial. In one embodiment the GAS AI-3 proteins are selected from the group consisting of SpyM3_0097, SpyM3_0098, SpyM3_0099, SpyM3_0100, SpyM3_0101, SpyM3_0102, SpyM3_0103, and SpyM3_0104. In another embodiment, the GAS AI-3 proteins are selected from the group consisting of SPs0099, SPs0100, SPs0101, SPs0102, SPs0103, SPs0104, SPs0105, and SPs0106. In yet another embodiment, the GAS AI-3 proteins are selected from the group consisting of orf77, orf78, orf79, orf80, orf81, orf82, orf83, and orf84. In a further embodiment, the GAS AI-3 proteins are selected from the group consisting of spyM18_0125, spyM18_0126, spyM18_0127, spyM18_0128, spyM18_0129, spyM18_0130, spyM18_0131, and spyM18_0132. In yet another embodiment the GAS AI-3 proteins are selected from the group consisting of SpyoM01000156, SpyoM01000155, SpyoM01000154, SpyoM01000153, SpyoM01000152, SpyoM01000151, SpyoM01000150, and SpyoM01000149.

The GAS AI-4 proteins may be selected from the group consisting of 19224133, 19224134, 19224135, 19224136, 19224137, 19224138, 19224139, 19224140, 19224141, 20010296_fimbrial, 20020069_fimbrial, CDC SS 635_fimbrial, ISS4883_fimbrial, and ISS4538_fimbrial. Preferably, the GAS-AI4 proteins are selected from the group consisting of 19224134, 19224135, 19224137, 19224139, 19224141, 20010296_fimbrial, 20020069_fimbrial, CDC SS 635_fimbrial, ISS4883_fimbrial, and ISS4538_fimbrial.

As yet another example, the invention includes an immunogenic composition comprising one or more of any combination of *S. pneumonaie* from TIGR4, *S. pneumonaie* strain 670, *S. pneumonaie* from 19A Hungary 6, *S. pneumonaie* from 6B Finland 12, *S. pneumonaie* from 6B Spain 2, *S. pneumonaie* from 9V Spain 3, *S. pneumonaie* from 14 CSR 10, *S. pneumonaie* from 19F Taiwan 14, *S. pneumonaie* from 23F Taiwan 15, or *S. pneumonaie* from 23F Poland 16 AI proteins. One or more of the AI proteins may be a surface protein. Such surface proteins may contain an LPXTG motif (such as LPXTG (SEQ ID NO: 122)) and may bind fibrinogen, fibronectin, or collagen. One or more of the AI proteins may be a sortase.

The *S. pneumonaie* from TIGR4 AI proteins may be selected from the group consisting of SP0462, SP0463, SP0464, SP0465, SP0466, SP0467, SP0468. Preferably, the *S. pneumonaie* from TIGR4 AI proteins include SP0462, SP0463, or SP0464.

The *S. pneumonaie* strain 670 AI proteins may be selected from the group consisting of Orf1_670, Orf3_670, Orf4_670, Orf5_670, Orf6_670, Orf7_670, and Orf8_670. Preferably, the *S. pneumonaie* strain 670 AI proteins include Orf3_670, Orf4_670, or Orf5_670.

The *S. pneumonaie* from 19A Hungary 6 AI proteins may be selected from the group consisting of ORF2_19AH, ORF3_19AH, ORF4_19AH, ORF5_19AH, ORF6_19AH, ORF7_19AH, or ORF8_19AH.

The *S. pneumonaie* from 6B Finland 12 AI proteins may be selected from the group consisting of ORF2_6BF, ORF3_6BF, ORF4_6BF, ORF5_6BF, ORF6_6BF, ORF7_6BF , or ORF8_6BF.

The *S. pneumonaie* from 6B Spain 2 AI proteins may be selected from the group consisting of ORF2_6BSP, ORF3_6BSP, ORF4_6BSP, ORF5_6BSP, ORF6_6BSP, ORF7_6BSP , or ORF8_6BSP.

The *S. pneumonaie* from 9V Spain 3 AI proteins may be selected from the group consisting of ORF2_9VSP, ORF3_9VSP, ORF4_9VSP, ORF5_9VSP, ORF6_9VSP, ORF7_9VSP , or ORF8_9VSP.

The *S. pneumonaie* from 14 CSR 10 AI proteins may be selected from the group consisting of ORF2_14CSR, ORF3_14CSR, ORF4_14CSR, ORF5_14CSR, ORF6_14CSR, ORF7_14CSR , or ORF8_14CSR.

The *S. pneumonaie* from 19F Taiwan 14 AI proteins may be selected from the group consisting of ORF2_19FTW, ORF3_19FTW, ORF4_19FTW, ORF5_19FTW, ORF6_19FTW, ORF7_19FTW , or ORF8_19FTW.

The *S. pneumonaie* from 23F Taiwan 15 AI proteins may be selected from the group consisting of ORF2_23FTW, ORF3_23FTW, ORF4_23FTW, ORF5_23FTW, ORF6_23FTW, ORF7_23FTW, or ORF8_23FTW.

The *S. pneumonaie* from 23F Poland 16 AI proteins may be selected from the group consisting of ORF2_23FP, ORF3_23FP, ORF4_23FP, ORF5_23FP, ORF6_23FP, ORF7_23FP , or ORF8_23FP.

Preferably, the Gram positive bacteria AI proteins included in the immunogenic compositions of the invention can provide protection across more than one serotype or strain isolate. For example, the immunogenic composition may comprise a first AI protein, wherein the amino acid sequence of said AI protein is at least 90% (*i.e.*, at least 90, 91, 92, 93, 94, 95, 96, 97, 98, 99 or 100%) homologous to the amino acid sequence of a second AI protein, and wherein said first AI protein and said second AI protein are derived from the genomes of different serotypes of a Gram positive bacteria. The first AI protein may also be homologous to the amino acid sequence of a third AI protein, such that the first AI protein, the second AI protein and the third AI protein are derived from the genomes of different serotypes of a Gram positive bacteria. The first AI protein may also be homologous to the amino acid sequence of a fourth AI protein, such that the first AI protein, the second AI protein and the third AI protein are derived from the genomes of different serotypes of a Gram positive bacteria.

For example, preferably, the GBS AI proteins included in the immunogenic compositions of the invention can provide protection across more than one GBS serotype or strain isolate. For example, the immunogenic composition may comprise a first GBS AI protein, wherein the amino acid sequence of said AI protein is at least 90% (*i.e.*, at least 90, 91, 92, 93, 94, 95, 96, 97, 98, 99 or 100%) homologous to the amino acid sequence of a second GBS AI protein, and wherein said first AI protein and said second AI protein are derived from the genomes of different GBS serotypes. The first GBS AI protein may also be homologous to the amino acid sequence of a third GBS AI protein, such that the first AI protein, the second AI protein and the third AI protein are derived from the genomes of different GBS serotypes. The first AI protein may also be homologous to the amino acid sequence of a fourth GBS AI protein, such that the first AI protein, the second AI protein and the third AI protein are derived from the genomes of different GBS serotypes.

The first AI protein may be selected from an AI-1 protein or an AI-2 protein. For example, the first AI protein may be a GBS AI-1 surface protein such as GBS 80. The amino acid sequence of GBS 80 from GBS serotype V, strain isolate 2603 is greater than 90% homologous to the GBS 80 amino acid sequence from GBS serotype III, strain isolates NEM316 and COH1 and the GBS 80 amino acid sequence from GBS serotype 1a, strain isolate A909.

As another example, the first AI protein may be GBS 104. The amino acid sequence of GBS 104 from GBS serotype V, strain isolate 2603 is greater than 90% homologous to the GBS 104 amino acid sequence from GBS serotype III, strain isolates NEM316 and COH1, the GBS 104 amino acid sequence from GBS serotype 1a, strain isolate A909, and the GBS 104 amino acid sequence serotype II, strain isolate 18RS21.

Table 12 provides the amino acid sequence identity of GBS 80 and GBS 104 across GBS serotypes Ia, Ib, II, III, V, and VIII. The GBS strains in which genes encoding GBS 80 and GBS 104 were identified share, on average, 99.88 and 99.96 amino acid sequence identity, respectively. This high degree of amino acid identity indicates that an immunogenic composition comprising a first protein of GBS 80 or GBS 104 may provide protection across more than one GBS serotype or strain isolate.

Table 12. Conservation of GBS 80 and GBS 104 amino acid sequences

Serotype	Strains	GBS 80		GBS 104	
		cGH	%AA identity	cGH	%AA identity
Ia	**090**	+	99.79	+	100.00
	A909	+	100.00	+	100.00
	515	-		-	
	DK1	-		-	
	DK8	-		-	
	Davis	-		-	
Ib	**7357b**	+	100.00	+	
	H36B	-		-	
II	**18RS21**	-		+	100.00
	DK21	-		-	
III	**NEM316**	+	100.00	+	100.00
	COH31	+	100.00	+	
	D136	+	100.00	+	
	M732	+	100.00	+	99.88
	COH1	+	99.79	+	99.88
	M781	+	99.79	+	99.88
No type	**CJB110**	+	99.37	+	100.00
	1169NT	-		-	
V	**CJB111**	+	100.00	+	100.00
	2603	+	100.00	+	100.00
VIII	**JM130013**	+	99.79	+	100.00
	SMU014	+	100.00	+	

Serotype	Strains	GBS 80		GBS 104	
		cGH	%AA identity	cGH	%AA identity
total		14/22	99.88+/-0.19	15/22	99.96 +/-0.056

As another example, the first AI protein may be an AI-2 protein such as GBS 67. The amino acid sequence of GBS 67 from GBS serotype V, strain isolate 2603 is greater than 90% homologous to the GBS 67 amino acid sequence from GBS serotype III, strain isolate NEM316, the GBS 67 amino acid sequence from GBS serotype 1b, strain isolate H36B, and the GBS 67 amino acid sequence from GBS serotype II, strain isolate 17RS21.

As another example, the first AI protein may be an AI-2 protein such as spb1. The amino acid sequence of spb1 from GBS serotype III, strain isolate COH1 is greater than 90% homologous to the spb1 amino acid sequence from GBS serotype Ia, strain isolate A909.

As yet another example, the first AI protein may be an AI-2 protein such as GBS 59. The amino acid sequence of GBS 59 from GBS serotype II, strain isolate 18RS21 is 100% homologous to the GBS 59 amino acid sequence from GBS serotype V, strain isolate 2603. The amino acid sequence of GBS 59 from GBS serotype V, strain isolate CJB111 is 98% homologous to the GBS 59 amino acid sequence from GBS serotype III, strain isolate NEM316.

The compositions of the invention may also be designed to include Gram positive AI proteins from divergent serotypes or strain isolates, *i.e.*, to include a first AI protein which is present in one collection of serotypes or strain isolates of a Gram positive bacteria and a second AI protein which is present in those serotypes or strain isolates not represented by the first AI protein.

For example, the invention may include an immunogenic composition comprising a first and second Gram positive bacteria AI protein, wherein a polynucleotide sequence encoding for the full length sequence of the first AI protein is not present in a similar Gram positive bacterial genome comprising a polynucleotide sequence encoding for the second AI protein.

The compositions of the invention may also be designed to include AI proteins from divergent GBS serotypes or strain isolates, *i.e.*, to include a first AI protein which is present in one collection of GBS serotypes or strain isolates and a second AI protein which is present in those serotypes or strain isolates not represented by the first AI protein.

For example, the invention may include an immunogenic composition comprising a first and second GBS AI protein, wherein a polynucleotide sequence encoding for the full length sequence of the first GBS AI protein is not present in a genome comprising a polynucleotide sequence encoding for the second GBS AI protein. For example, the first AI protein could be GBS 80 (such as the GBS 80 sequence from GBS serotype V, strain isolate 2603). As previously discussed (and depicted in Figure 2), the sequence for GBS 80 in GBS sertoype II, strain isolate 18RS21 is disrupted. In this instance, the second AI protein could be GBS 104 or GBS 67 (sequences selected from the GBS serotype II, strain isolate 18RS21).

Further, the the invention may include an immunogenic composition comprising a first and second GBS AI protein, wherein the first GBS AI protein has detectable surface exposure on a first GBS strain or serotype but not a second GBS strain or serotype and the second GBS AI protein has detectable surface exposure on a second GBS strain or serotype but not a first GBS strain or serotype. For example, the first AI protein could be GBS 80 and the second AI protein could be GBS 67. As seen in Table 15, there are some GBS serotypes and strains that have surface exposed GBS 80 but that do not have surface exposed GBS 67 and vice versa. An immunogenic composition comprising a GBS 80 and a GBS 67 protein may provide protection across a wider group of GBS strains and serotypes.

Table 15: Antigen surface exposure of GBS 80, GBS 322, GBS 104, and GBS 67

GBS strains	Type	GBS 80	GBS 322	GBS 104	GBS 67
DK1*	Ia	0	nd	237	478
DK8*		0	213	151	475
Davis*		0	86	271	430
515*		0	227	262	409
090		0	0	0	0
A909		0	0	0	0
2986		0	0	157	397
5551		0	36	384	485
2177	Ib	477	323	328	66
H36B*		0	105	518	444
7357b-		91	102	309	316
2129		57	71	132	0
5518		31	nd	60	28
COH1	III	305	130	305	0
D136C		16	460	226	406
COH31		0	479	71	273
M732		105	292	101	0
M781		65	224	136	0
1998		95	288	205	350
5376		165	76	156	0
5435		93	88	100	0
18RS21	II	0	471	50	103
DK21*		0	342	419	331
3050		43	188	289	460
5401		170	135	494	618
2141		0	76	0	69
CJB111	V	365	58	355	481
2603		62	293	100	105
5364		454	463	379	394
2110		0	11	345	589
2274	IV	113	161	465	484
1999		0	55	492	453
2210		0	0	363	574
2928	VII	0	0	0	0
SMU071	VIII	556	170	393	79
JM9130013		587	133	436	83
2189		0	0	0	0
5408		0	0	159	433
CJB110	NT	71	587	169	245
1169*		0	213	371	443
Δ Mean > 100		9/40	22/38	32/40	25/40
		22%	58%	80%	62%

Alternatively, the invention may include an immunogenic composition comprising a first and second Gram positive bacteria AI protein, wherein the polynucleotide sequence encoding the sequence of the first AI protein is less than 90 % (*i.e.*, less than 90, 88, 86, 84, 82, 80, 78, 76, 74, 72, 70, 65, 60, 55, 50, 45, 40, 35 or 30 percent) homologous than the corresponding sequence in the genome of the second AI protein.

The invention may include an immunogenic composition comprising a first and second GBS AI protein, wherein the polynucleotide sequence encoding the sequence of the first GBS AI protein is less than 90 % (*i.e.*, less than 90, 88, 86, 84, 82, 80, 78, 76, 74, 72, 70, 65, 60, 55, 50, 45, 40, 35 or 30 percent) homologous than the corresponding sequence in the genome of the second GBS AI protein. For example, the first GBS AI protein could be GBS 67 (such as the GBS 67 sequence from GBS serotype 1b, strain isolate H36B). As shown in Figures 2 and 4, the GBS 67 sequence for this strain is less than 90% homologous (87%) to the corresponding GBS 67 sequence in GBS serotype V, strain isolate 2603. In this instance, the second GBS AI protein could then be the GBS 80 sequence from GBS serotype V, strain isolate 2603.

An example immunogenic composition of the invention may comprise adhesin island proteins GBS 80, GBS 104, GBS 67, and GBS 59, and non-AI protein GBS 322. FACS analysis of different GBS strains demonstrates that at least one of these five proteins is always found to be expressed on the surface of GBS bacteria. An initial FACS analysis of 70 strains of GBS bacteria, obtained from the CDC in the United States (33 strains), ISS in Italy (17 strains), and Houston/Harvard (20 strains), detected surface exposure of at least one of GBS 80, GBS 104, GBS 322, GBS 67, or GBS 59 on the surface of the GBS bacteria. Figure 227 provides the FACS data obtained for surface exposure of GBS 80, GBS 104, GBS 67, GBS 322, and GBS 59 on each of 37 GBS strains. Figure 228 provides the FACS data obtained for surface exposure of GBS 80, GBS 104, GBS 67, GBS 322, and GBS 59 on each of 41 GBS strains obtained from the CDC. As can be seen from Figures 227 and 228, each GBS strain had surface expression of at least one of GBS 80, GBS 104, GBS 67, GBS 322, and GBS 59. The surface exposure of at least one of these proteins on each bacterial strain indicates that an immunogenic composition comprising these proteins will provide wide protection across GBS strains and serotypes.

The surface exposed GBS 80, GBS 104, GBS 67, GBS 322, and GBS 59 proteins are also present at high levels as determined by FACS. Table 49 summarizes the FACS results for the initial 70 GBS strains examined for GBS 80, GBS 104, GBS 67, GBS 322, and GBS 59 surface expression. A protein was designated as having high levels of surface expression of a protein if a five-fold shift in fluorescence was observed when using antibodies for the protein relative to preimmune control serum.

Table 49: Exposure Levels of GBS 80, GBS 104, GBS 67, GBS 322, and GBS 59 on GBS Strains

5-fold shift in fluorescence by FACS	GBS 80	GBS 104	GBS 67	GBS 59	GBS 322
	17/70	14/70	49/70	46/70	33/70
	24%	20%	70%	66%	47%

Table 50 details which of the surface proteins is highly expressed on the different GBS serotype.

Table 50: High Levels of Surface Protein Expression on GBS Serotypes

5-fold shift in fluorescence by FACS	GBS 80	GBS 104	GBS 67	GBS 59	GBS 322
Ia + Ib + III	4/36	2/36	22/36	20/36	18/36
II + V	11/25	9/25	21/25	21/25	13/25
Others	2/9	3/9	6/9	5/9	2/9

Alternatively, the immunogenic composition of the invention may include GBS 80, GBS 104, GBS 67, and GBS 322. Assuming that protein antigens that are highly accessible to antibodies confer 100% protection with suitable adjuvants, an immunogenic composition containing GBS 80, GBS 104, GBS 67, GBS 59 and GBS 322 will provide protection for 89% of GBS strains and serotypes, the same percentage as an immunogenic composition containing GBS 80, GBS 104, GBS 67, and GBS 322 proteins. See Figure 229. However, it may be preferable to include GBS 59 in the composition to increase its immunogenic strength. As seen from Table 50, GBS 59 is highly expressed on the surface two-thirds of GBS bacteria examined by FACS analysis, unlike GBS 80, GBS 104, and GBS 322, which are highly expressed in less than half of GBS bacteria examined. GBS 59 opsonophagocytic activity is also comparable to that of a mix of GBS 322, GBS 104, GBS 67, and GBS 80 proteins. See Figure 230.

By way of another example, preferably, the GAS AI proteins included in the immunogenic compositions of the invention can provide protection across more than one GAS serotype or strain isolate. For example, the immunogenic composition may comprise a first GAS AI protein, wherein the amino acid sequence of said AI protein is at least 90% (*i.e.*, at least 90, 91, 92, 93, 94, 95, 96, 97, 98, 99 or 100%) homologous to the amino acid sequence of a second GAS AI protein, and wherein said first AI protein and said second AI protein are derived from the genomes of different GAS serotypes. The first GAS AI protein may also be homologous to the amino acid sequence of a third GAS AI protein, such that the first AI protein, the second AI protein and the third AI protein are derived from the genomes of different GAS serotypes. The first AI protein may also be homologous to the amino acid sequence of a fourth GAS AI protein, such that the first AI protein, the second AI protein and the third AI protein are derived from the genomes of different GAS serotypes.

The compositions of the invention may also be designed to include GAS AI proteins from divergent serotypes or strain isolates, *i.e.*, to include a first AI protein which is present in one collection of serotypes or strain isolates of a GAS bacteria and a second AI protein which is present in those serotypes or strain isolates not represented by the first AI protein.

For example, the first AI protein could be a prtF2 protein (such as the 19224141 protein from GAS serotype M12, strain isolate A735). As previously discussed (and depicted in Figure 164), the sequence for a prtF2 protein is not present in GAS AI types 1 or 2. In this instance, the second AI protein could be collagen binding protein M6_Spy0159 (from M6 isolate (MGAS10394), which comprises an AI-1) or GAS15 (from M1 isolate (SF370), which comprises an AI-2).

Further, the invention may include an immunogenic composition comprising a first and second GAS AI protein, wherein the first GAS AI protein has detectable surface exposure on a first GAS strain or serotype but not a second GAS strain or serotype and the second GAS AI protein has detectable surface exposure on a second GAS strain or serotype but not a first GAS strain or serotype.

The invention may include an immunogenic composition comprising a first and second GAS AI protein, wherein the polynucleotide sequence encoding the sequence of the first GAS AI protein is less than 90 % (*i.e.*, less than 90, 88, 86, 84, 82, 80, 78, 76, 74, 72, 70, 65, 60, 55, 50, 45, 40, 35 or 30 percent) homologous than the corresponding sequence in the genome of the second GAS AI protein. Preferably the first and second GAS AI proteins are subunits of the pilus. More preferably the first and second GAS AI proteins are selected from the major pilus forming proteins (i.e., M6_Spy0160 from M6 strain 10394, SPy0128 from M1 strain SF370, SpyM3_0100 from M3 strain 315, SPs0102 from M3 strain SSI, orf80 from M5 isolate Manfredo, spyM18_0128 from M18 strain 8232, SpyoM01000153 from M49 strain 591, 19224137 from M12 strain A735, fimbrial structural subunit from M77 strain ISS4959, fimbrial structural subunit from M44 strain ISS3776, fimbrial structural subunit from M50 strain ISS3776 ISS 4538, fimbrial structural subunit from M12strain CDC SS635, fimbrial structural subunit from M23 strain DSM2071, fimbrial structural subunit from M6 strain CDC SS410). Table 45 provides the percent identity between the amino acidic sequences of each of the main pilus forming subunits from GAS AI-1, AI-2, AI-3, and AI-4 representative strains (i.e., M6_Spy0160 from M6 strain 10394, SPy0128 from M1 strain SF370, SpyM3_0100 from M3 strain 315, SPs0102 from M3 strain SSI, orf80 from M5 isolate Manfredo, spyM18_0128 from M18 strain 8232, SpyoM01000153 from M49 strain 591, 19224137 from M12 strain A735, Fimbrial structural subunit from M77 strain ISS4959, fimbrial structural subunit from M44 strain ISS3776, fimbrial structural subunit from M50 strain ISS3776 ISS 4538, fimbrial structural subunit from M12strain CDC SS635, fimbrial structural subunit from M23 strain DSM2071, fimbrial structural subunit from M6 strain CDC SS410).

Table 45: Comparison of Amino Acid Sequences of Major Pilus Proteins in the Four GAS AIs

		AI-1	**AI-2**	**AI-3**			**AI-4**
		M6-10394	M1-370	M3-315	M5-Manfredo	M18-8232	M12-A735
AI-1	M6-10394	**100%** M6-10394 M6-SS-410 **56%** M23-DSM2071	23%	25%	23%	24%	26%
AI-2	M1-370	23%	**100%**	40%	41%	38%	40%
AI-3	M3-315	25%	40%	**100%** M3-315 M3-SSI-1	64%	67%	61%
	M5-Manfredo	23%	39%	64%	**100%** M5-Manfredo M44-3776 M77-4959	60%	65%

	M18-8232	24%	38%	67%	60%	**90%** M18-8232 M49-591	62%
AI-4	M12-A735	26%	40%	61%	65%	62%	**97-100%** M12-A735 M12-SS-635 M50-4538

For example, the first main pilus subunit may be selected from bacteria of GAS serotype M6 strain 10394 and the second main pilus subunit may be selected from bacteria of GAS serotype M1 strain 370. As can be seen from Table 45, the main pilus subunits encoded by these strains of bacteria share only 23% nucleotide identity. An immunogenic composition comprising pilus main subunits from each of these strains of bacteria is expected to provide protection across a wider group of GAS strains and serotypes. Other examples of main pilus subunits that can be used in combination to provide increased protection across a wider range of GAS strains and serotypes include proteins encoded by GAS serotype M5 Manfredo isolate and serotype M6 strain 10394, which share 23% sequence identity, GAS serotype M18 strain 8232 and serotype M1 strain 370, which share 38% sequence identity, GAS serotype M3 strain 315 and serotype M12 strain A735, which share 61% sequence identity, and GAS serotype M3 strain 315 and serotype M6 strain 10394 which share 25% sequence identity.

As also can be seen from Table 45, the amino acid sequences of the four types of main pilus subunits present in GAS are relatively divergent. Figures 198-201 provide further tables comparing the percent identity of adhesin island-encoded surface exposed proteins for different GAS serotypes relative to other GAS serotypes harbouring an adhesin island of the same or a different subtype (GAS AI-1, GAS AI-2, GAS AI-3, and GAS AI-4). See also further discussion below.

Immunizations with the Adhesin Island proteins of the invention are discussed further in the Examples.

Co-expression of GBS Adhesin Island proteins and role of GBS AI proteins in surface presentation

In addition to the use of the GBS adhesin island proteins for cross strain and cross serotype protection, Applicants have identified interactions between adhesin island proteins which appear to affect the delivery or presentation of the surface proteins on the surface of the bacteria.

In particular, Applicants have discovered that surface exposure of GBS 104 is dependent on the concurrent expression of GBS 80. As discussed further in Example 2, reverse transcriptase PCR analysis of AI-1 shows that all of the AI genes are co-transcribed as an operon. Applicants constructed a series of mutant GBS containing in frame deletions of various AI-1 genes. (A schematic of the GBS mutants is presented in Figure 7). FACS analysis of the various mutants comparing mean shift values using anti-GBS 80 versus anti-GBS 104 antibodies is presented in Figure 8. Removal of the GBS 80 operon prevented surface exposure of GBS 104; removal of the

GBS 104 operon did not affect surface exposure of GBS 80. While not being limited to a specific theory, it is thought that GBS 80 is involved in the transport or localization of GBS 104 to the surface of the bacteria. The two proteins may be oligomerized or otherwise associated. It is possible that this association involves a conformational change in GBS 104 that facilitates its transition to the surface of the GBS bacteria.

Pili structures that comprise GBS 104 appear to be of a lower molecular weight than pili structures lacking GBS 104. Figure 68 shows that polyclonal anti-GBS 104 antibodies (see lane marked α-104 POLIC.) cross-hybridize with smaller structures than do polyclonal anti-GBS 80 antibodies (see lane marked α-GBS 80 POLIC.).

In addition, Applicants have shown that removal of GBS 80 can cause attenuation, further suggesting the protein contributes to virulence. As described in more detail in Example 3, the LD_{50}'s for the Δ80 mutant and the Δ80, Δ104 double mutant were reduced by an order of magnitude compared to wildtype and Δ104 mutant.

The sortases within the adhesin island also appear to play a role in localization and presentation of the surface proteins. As discussed further in Example 4, FACS analysis of various sortase deletion mutants showed that removal of sortase SAG0648 prevented GBS 104 from reaching the surface and slightly reduced the surface exposure of GBS 80. When sortase SAG0647 and sortase SAG0648 were both knocked out, neither GBS 80 nor GBS 104 were surface exposed. Expression of either sortase alone was sufficient for GBS 80 to arrive at the bacterial surface. Expression of SAG0648, however, was required for GBS 104 surface localization.

Accordingly, the compositions of the invention may include two or more AI proteins, wherein the AI proteins are physically or chemically associated. For example, the two AI proteins may form an oligomer. In one embodiment, the associated proteins are two AI surface proteins, such as GBS 80 and GBS 104. The associated proteins may be AI surface proteins from different adhesin islands, including host cell adhesin island proteins if the AI surface proteins are expressed in a recombinant system. For example, the associated proteins may be GBS 80 and GBS 67.

Adhesin Island proteins from other Gram positive bacteria

Applicants' identification and analysis of the GBS adhesin islands and the immunological and biological functions of these AI proteins and their pilus structures provides insight into similar structures in other Gram positive bacteria.

As discussed above, "Adhesin Island" or "AI" refers to a series of open reading frames within a bacterial genome that encode for a collection of surface proteins and sortases. An Adhesin Island may encode for amino acid sequences comprising at least one surface protein. The Adhesin Island may encode at least one surface protein. Alternatively, an Adhesin Island may encode for at least two surface proteins and at least one sortase. Preferably, an Adhesin Island encodes for at least three surface proteins and at least two sortases. One or more of the surface proteins may include an LPXTG motif (such as LPXTG (SEQ ID NO: 122)) or other sortase substrate motif. One or more AI surface proteins may participate in the formation of a pilus structure on the surface of the Gram positive bacteria.

Gram positive adhesin islands of the invention preferably include a divergently transcribed transcriptional regulator. The transcriptional regulator may regulate the expression of the AI operon.

The invention includes a composition comprising one or more Gram positive bacteria AI surface proteins. Such AI surface proteins may be associated in an oligomeric or hyperoligomeric structure.

Preferred Gram positive adhesin island proteins for use in the invention may be derived from *Staphylococcus* (such as *S. aureus*), *Streptococcus* (such as *S. agalactiae* (GBS), *S. pyogenes* (GAS), *S. pneumonaie*, *S. mutans*), *Enterococcus* (such as *E. faecalis* and *E. faecium*), *Clostridium* (such as *C. difficile*), *Listeria* (such as *L. monocytogenes*) and *Corynebacterium* (such as *C. diphtheria*).

One or more of the Gram positive AI surface protein sequences typically include an LPXTG motif or other sortase substrate motif. Gram positive AI surface proteins of the invention may affect the ability of the Gram positive bacteria to adhere to and invade epithelial cells. AI surface proteins may also affect the ability of Gram positive bacteria to translocate through an epithelial cell layer. Preferably, one or more AI surface proteins are capable of binding to or otherwise associating with an epithelial cell surface. Gram positive AI surface proteins may also be able to bind to or associate with fibrinogen, fibronectin, or collagen.

Gram positive AI sortase proteins are predicted to be involved in the secretion and anchoring of the LPXTG containing surface proteins. A Gram positive bacteria AI may encode for at least one surface exposed protein. The Adhesin Island may encode at least one surface protein. Alternatively, a Gram positive bacteria AI may encode for at least two surface exposed proteins and at least one sortase. Preferably, a Gram positive AI encodes for at least three surface exposed proteins and at least two sortases.

Gram positive AI surface proteins may be covalently attached to the bacterial cell wall by membrane-associated transpeptidases, such as an AI sortase. The sortase may function to cleave the surface protein, preferably between the threonine and glycine residues of an LPXTG motif. The sortase may then assist in the formation of an amide link between the threonine carboxyl group and a cell wall precursor such as lipid II. The precursor can then be incorporated into the peptidoglycan via the transglycoslylation and transpeptidation reactions of bacterial wall synthesis. *See* Comfort et al., Infection & Immunity (2004) 72(5): 2710 – 2722. Typically, Gram positive bacteria AI surface proteins of the invention will contain an N-terminal leader or secretion signal to facilitate translocation of the surface protein across the bacterial membrane.

Gram positive bacteria AI surface proteins of the invention may affect the ability of the Gram positive bacteria to adhere to and invade target host cells, such as epithelial cells. Gram positive bacteria AI surface proteins may also affect the ability of the gram positive bacteria to translocate through an epithelial cell layer. Preferably, one or more of the Gram positive AI surface proteins are capable of binding to or other associating with an epithelial cell surface. Further, one or more Gram positive AI surface proteins may bind to fibrinogen, fibronectin, or collagen protein.

In one embodiment, the invention includes a composition comprising oligomeric, pilus-like structures comprising a Gram positive bacteria AI surface protein. The oligomeric, pilus-like structure may comprise numerous units of the AI surface protein. Preferably, the oligomeric, pilus-like structures comprise two or more AI surface proteins. Still more preferably, the oligomeric, pilus-like structure comprises a hyper-oligomeric pilus-like structure comprising at least two (*e.g.*, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 20, 25, 30, 35, 40, 45, 50, 60, 70, 80, 90, 100, 120, 140, 150, 200 or more) oligomeric subunits, wherein each subunit comprises an AI surface protein or a fragment thereof. The oligomeric subunits may be covalently associated via a conserved lysine within a pilin motif. The oligomeric subunits may be covalently associated via an LPXTG motif, preferably, via the threonine amino acid residue.

Gram positive bacteria AI surface proteins or fragments thereof to be incorporated into the oligomeric, pilus-like structures of the invention will preferably include one or both of a pilin motif comprising a conserved lysine residue and an E box motif comprising a conserved glutamic acid residue.

The oligomeric, pilus like structures may be used alone or in the combinations of the invention. In one embodiment, the invention comprises a Gram positive bacteria Adhesin Island in oligomeric form, preferably in a hyperoligomeric form.

The oligomeric, pilus-like structures of the invention may be combined with one or more additional Gram positive AI proteins (from the same or a different Gram positive species or genus). In one embodiment, the oligomeric, pilus-like structures comprise one or more Gram positive bacteria AI surface proteins in combination with a second Gram positive bacteria protein. The second Gram positive bacteria protein may be a known antigen, and need not normally be associated with an AI protein.

The oligomeric, pilus-like structures may be isolated or purified from bacterial cultures overexpressing a Gram positive bacteria AI surface protein. The invention therefore includes a method for manufacturing an oligomeric Adhesin Island surface antigen comprising culturing a Gram positive bacteria adapted for increased AI protein expression and isolation of the expressed oligomeric Adhesin Island protein from the Gram positive bacteria. The AI protein may be collected from secretions into the supernatant or it may be purified from the bacterial surface. The method may further comprise purification of the expressed Adhesin Island protein. Preferably, the Adhesin Island protein is in a hyperoligomeric form.

Gram positive bacteria are preferably adapted to increase AI protein expression by at least two (*e.g.*, 2, 3, 4, 5, 8, 10, 15, 20, 25, 30, 35, 40, 45, 50, 60, 70, 80, 90, 100, 125, 150 or 200) times wild type expression levels.

Gram positive bacteria may be adapted to increase AI protein expression by means known in the art, including methods of increasing gene dosage and methods of gene upregulation. Such means include, for example, transformation of the Gram positive bacteria with a plasmid encoding the AI protein. The plasmid may include a strong promoter or it may include multiple copies of the sequence encoding the AI protein. Optionally, the sequence encoding the AI protein within the Gram positive bacterial genome may be deleted. Alternatively, or in addition, the promoter regulating the Gram positive Adhesin Island may be modified to increase expression.

The invention further includes Gram positive bacteria which have been adapted to produce increased levels of AI surface protein. In particular, the invention includes Gram positive bacteria which have been adapted to produce oligomeric or hyperoligomeric AI surface protein. In one embodiment, the Gram positive bacteria of the invention are inactivated or attenuated to permit *in vivo* delivery of the whole bacteria, with the AI surface protein exposed on its surface.

The invention further includes Gram positive bacteria which have been adapted to have increased levels of expressed AI protein incorporated in pili on their surface. The Gram positive bacteria may be adapted to have increased exposure of oligomeric or hyperoligomeric AI proteins on its surface by increasing expression levels of a signal peptidase polypeptide. Increased levels of a local signal peptidase expression in Gram positive bacteria (such us LepA in GAS) are expected to result in increased exposure of pili proteins on the surface of Gram positive bacteria. Increased expression of a leader peptidase in Gram positive may be achieved by any means known in the art, such as increasing gene dosage and methods of gene upregulation. The Gram positive bacteria adapted to have increased levels of leader peptidase may additionally be adapted to express increased levels of at least one pili protein.

Alternatively, the AI proteins of the invention may be expressed on the surface of a non-pathogenic Gram positive bacteria, such as *Streptococus gordonii* (See, e.g., Byrd et al., "Biological consequences of antigen and cytokine co-expression by recombinant *Streptococcus gordonii* vaccine vectors", Vaccine (2002) 20:2197-2205) or *Lactococcus lactis* (See, e.g., Mannam et al., "Mucosal VaccineMade from Live, Recombinant Lactococcus lactis Protects Mice against Pharangeal Infection with *Streptococcus pyogenes*" Infection and Immunity (2004) 72(6):3444-3450). It has already been demonstrated, above, that *L. lactis* expresses GBS and GAS AI polypeptides in oligomeric form and on its surface.

Alternatively, the oligomeric, pilus-like structures may be produced recombinantly. If produced in a recombinant host cell system, the Gram positive bacteria AI surface protein will preferably be expressed in coordination with the expression of one or more of the AI sortases of the invention. Such AI sortases will facilitate oligomeric or hyperoligomeric formation of the AI surface protein subunits.

Gram positive AI Sortases of the invention will typically have a signal peptide sequence within the first 70 amino acid residues. They may also include a transmembrane sequence within 50 amino acid residues of the C terminus. The sortases may also include at least one basic amino acid residue within the last 8 amino acids. Preferably, the sortases have one or more active site residues, such as a catalytic cysteine and histidine.

Adhesin island surface proteins from two or more Gram positive bacterial genus or species may be combined to provide an immunogenic composition for prophylactic or therapeutic treatment of disease or infection of two more Gram positive bacterial genus or species. Optionally, the adhesin island surface proteins may be associated together in an oligomeric or hyperoligomeric structure.

In one embodiment, the invention comprises an adhesin island surface proteins from two or more *Streptococcus* species. For example, the invention includes a composition comprising a GBS AI surface protein and a GAS adhesin island surface protein. As another example, the invention includes a composition comprising a GAS adhesin island surface protein and a *S. pneumoniae* adhesin island surface protein.

In one embodiment, the invention comprises an adhesin island surface protein from two or more Gram positive bacterial genus. For example, the invention includes a composition comprising a *Streptococcus* adhesin island protein and a *Corynebacterium* adhesin island protein.

Examples of AI sequences in several Gram positive bacteria are discussed further below.

Streptococcus pyogenes (GAS)

As discussed above, Applicants have identified at least four different GAS Adhesin Islands. These adhesion islands are thought to encode surface proteins which are important in the bacteria's virulence, and Applicants have obtained the first electron micrographs revealing the presence of these adhesin island proteins in hyperoligomeric pilus structures on the surface of Group A Streptococcus.

Group A Streptococcus is a human specific pathogen which causes a wide variety of diseases ranging from pharyngitis and impetigo through life threatening invasive disease and necrotizing fasciitis. In addition, post-streptococcal autoimmune responses are still a major cause of cardiac pathology in children.

Group A Streptococcal infection of its human host can generally occur in three phases. The first phase involves attachment and/or invasion of the bacteria into host tissue and multiplication of the bacteria within the extracellular spaces. Generally this attachment phase begins in the throat or the skin. The deeper the tissue level infected, the more severe the damage that can be caused. In the second stage of infection, the bacteria secrete a soluble toxin that diffuses into the surrounding tissue or even systemically through the vasculature. This toxin binds to susceptible host cell receptors and triggers innappropropriate immune responses by these host cells, resulting in pathology. Because the toxin can diffuse throughout the host, the necrosis directly caused by the GAS toxins may be physically located in sites distant from the bacterial infection. The final phase of GAS infection can occur long after the original bacteria have been cleared from the host system. At this stage, the host's previous immune response to the GAS bacteria due to cross reactivity between epitopes of a GAS surface protein, M, and host tissues, such as the heart. A general review of GAS infection can be found in Principles of Bacterial Pathogeneis, Groisman ed., Chapter 15 (2001).

In order to prevent the pathogenic effects associated with the later stages of GAS infection, an effective vaccine against GAS will preferably facilitate host elimination of the bacteria during the initial attachment and invasion stage.

Isolates of Group A Streptococcus are historically classified according to the M surface protein described above. The M protein is surface exposed trypsin-sensitive protein generally comprising two polypeptide chains complexed in an alpha helical formation. The carboxyl terminus is anchored in the cytoplasmic membrane and is highly conserved among all group A streptococci. The amino terminus, which extends through the cell wall to the cell surface, is responsible for the antigenic variability observed among the 80 or more serotypes of M proteins.

A second layer of classification is based on a variable, trypsin-resistant surface antigen, commonly referred to as the T-antigen. Decades of epidemiology based on M and T serological typing have been central to studies on the biological diversity and disease causing potential of Group A Streptococci. While the M-protein component and its inherent variability have been extensively characterized, even after five decades of study, there is still very little known about the structure and variability of T-antigens. Antisera to define T types are commercially available from several sources, including Sevapharma (http://www.sevapharma.cz/en).

The gene coding for one form of T-antigen, T-type 6, from an M6 strain of GAS (D741) has been cloned and characterized and maps to an approximately 11 kb highly variable pathogenicity island. Schneewind et al., J Bacteriol. (1990) 172(6):3310 – 3317. This island is known as the Fibronectin-binding, Collagen-binding T-antigen (FCT) region because it contains, in addition to the T6 coding gene (*tee6*), members of a family of genes coding for Extra Cellular Matrix (ECM) binding proteins. Bessen et al., Infection & Immunity (2002) 70(3):1159-1167. Several of the protein products of this gene family have been shown to directly bind either fibronectin and/or collagen. *See* Hanski et al., Infection & Immunity (1992) 60(12):5119-5125; Talay et al., Infection & Immunity (1992(60(9):3837-3844; Jaffe et al. (1996) 21(2):373-384; Rocha et al., Adv Exp Med Biol. (1997) 418:737-739; Kreikemeyer et al., J Biol Chem (2004) 279(16):15850-15859; Podbielski et al., Mol. Microbiol. (1999) 31(4):1051-64; and Kreikemeyer et al., Int. J. Med Microbiol (2004) 294(2-3):177-88. In some cases direct evidence for a role of these proteins in adhesion and invasion has been obtained.

Applicants raised antiserum against a recombinant product of the *tee6* gene and used it to explore the expression of T6 in M6 strain ISS3650. In immunoblot of mutanolysin extracts of this strain, the antiserum recognized, in addition to a band corresponding to the predicted molecular mass of the *tee6* gene product, very high molecular weight ladders ranging in mobility from about 100 kDa to beyond the resolution of the 3-8% gradient gels used. See Figure 163A, last lane labeled "M6_Tee6."

This pattern of high molecular weight products is similar to that observed in immunoblots of the protein components of the pili identified in *Streptococcus agalactiae* (described above) and previously in *Corynebacterium diphtheriae*. Electron microscropy of strain M6 ISS3650 with antisera specific for the product of *tee6* revealed abundant surface staining and long pilus like structures extending up to 700 nanometers from the bacterial surface, revealing that the T6 protein, one of the antigens recognized in the original Lancefield serotyping system, is located within a GAS Adhesin Island (GAS AI-1) and forms long covalently linked pilus structures. See Figure 163I.

In addition to the *tee6* gene, the FCT region in M6_ISS3650 (GAS AI-1) contains two other genes (*prtF1* and *cpa*) predicted to code for surface exposed proteins; these proteins are characterized as containing the cell wall attachment motif LPXTG. Western blot analysis using antiserum specific for PrtF1 detected a single molecular species with electrophoretic mobility corresponding to the predicted molecular mass of the protein and one smaller band of unknown origin. Western blot analysis using antisera specific for Cpa recognized a high molecular weight covalently linked ladder (Fig 163A, second lane). Immunogold labelling of Cpa with specific antiserum followed by transmission electron microscopy detected an abundance of Cpa at the cell surface and only occasional structures extending from the cell surface (Fig. 163J).

Four classes of FCT region can be discerned by the types and order of the genes contained within the region. The FCT region of strains of types M3, M5, M18 and M49 have a similar organization whereas those of M6, M1 and M12 differ. See Figure 164. As discussed below, these four FCT regions correlate to four GAS Adhesin Island types (AI-1, AI-2, AI-3 and AI-4).

Applicants discovery of genes coding for pili in the FCT region of strain M6_ISS3650 prompted them to examine the predicted surface exposed proteins in the variant FCT regions of three other GAS strains of having different M-type (M1_SF370, M5_ISS4883 and M12_20010296) representing the other three FCT variants. Each gene present in the FCT region of each bacteria was cloned and expressed. Antisera specific for each recombinant protein was then used to probe mutanolysin extracts of the respective strains (*6*). In M1 strain SF370, there are three predicted surface proteins (Cpa (also referred to as M1_126 and GAS 15), M1_128 (a fimbrial protein also referred to as Spy0128 and GAS 16), and M1_130 (also referred to as Spy0130 and GAS 18)) (GAS AI-2). Antisera specific for each surface protein reacted with a ladder of high molecular weight material (Fig. 163B). Immunogold staining of M1 strain SF370 with antiserum specific for M1_128 revealed pili structures similar to those seen when M6 strain ISS3650 was immunogold stained with antiserum specific for teе6 (See Fig 1163K). Antisera specific for surface proteins Cpa and M1_130 revealed abundant surface staining and occasional structures extending from the surface of M1 strain SF370 bacteria (Fig. 163S).

The M1_128 protein appears to be necessary for polymerization of Cpa and M1_130 proteins. If the M1_128 gene in M1_SF370 was deleted, Western blot analysis using antibodies that hybridize to Cpa and M1_130 no longer detected high molecular weight ladders comprising the Cpa and M1_130 proteins (Fig. 163 E). See also Figures 177 A-C which provide the results of Western blot analysis of the M1_128 (Δ128) deleted bacteria using anti-M1_130 antiserum (Figure 177 A), anti-M1_128 antiserum (Figure 177 B), and anti-M1_126 antiserum (Figure 177

C). High molecular weight ladders, indicative of pilus formation on the surface of M1 strain SF370, could not be detected by any of the three antisera in Δ128 bacteria. If the Δ128 bacteria were transformed with a plasmid containing the gene for M1_128, Western blot analysis using antisera specific for Cpa and M1_130 again detected high molecular weight ladders (Figure 163 H).

In agreement with the Western blot analysis, immunoelectron microscopy failed to detect pilus assembly on the Δ128 strain SF370 bacteria using M1_128 antisera (Figure 178 B). Although Δ128 SF370 bacteria were unable to form pili, M1_126 (cpa) and M1_130, which contain sortase substrate motifs, were present on the bacteria's surface. FACS analysis of the M1_128 deleted (Δ128) strain SF370 bacteria also detected both M1_126 and M1_130 on the surface of the Δ128 strain SF370 bacteria. See Figure 179 D and F, which show a shift in fluorescence when antibodies immunoreactive to M1_126 and M1_130 are used on Δ128 bacteria. As expected, virtually no shift in fluorescence is observed when antibodies immunoreactive to M1_128 are used with the Δ128 bacteria (Figure 179 E).

By contrast, deletion of the M1_130 gene did not effect polymerization of M1_128 (Fig. 163 F). See also Figures 177 A-C, which provide Western blot analysis results of the M1_130 deleted (Δ130) strain SF370 bacteria using anti-M1_130 (Figure 177 A), anti-M1_128 (Figure 177 B), and anti-M1_126 antiserum (Figure 177 C). The anti-M1_128 and anti-M1_126 antiserum both detected the presence of high molecular weight ladders in the Δ130 strain SF370 bacteria, indicating that the Δ130 bacteria form pili that comprise M1_126 and M1_128 polypeptides in the absence of M1_130. As expected, the Western blot probed with antiserum immunoreactive with M1_130 did not detect any proteins for the Δ130 bacteria(Figure 177A).

Hence, the composition of the pili in GAS resembles that previously described for both *C. diphtheria* (*7, 8*) and *S. agalactiae* (described above) (*9*) in that each pilus is formed by a backbone component which abundantly stains the pili in EM and is essential for the incorporation of the other components.

Also similar to *C. diphtheria*, elimination of the srtC1 gene from the FCT region of M1_SF370 abolished polymerization of all three proteins and assembly of pili (Fig. 163 G). See also Figures 177 A-C, which provide Western blot analysis of the SrtC1 deleted (ΔSrtC1) strain SF370 bacteria using anti-M1_130 (Figure 177 A), anti-M1_128 (Figure 177 B), and anti-M1_126 antiserum (Figure 177 C). None of the three antisera immunoreacted with high molecular weight structures (pili) in the ΔSrtC1 bacteria. Confirming that deletion of the SrtC1 gene abrogates pilus assembly in strain SF370, immunoelectron microscopy using antisera against M1_128 failed to detect pilus formation on the bacteria surface. See Figure 178 C. Although no assembled pili were detected on ΔSrtC1 SF370, M1_128 proteins could be detected on the surface of SF370. Thus, it appeared that SrtC1 deletion prevented pilus assembly on the surface of the SF370 bacteria, but not anchoring of the proteins that comprise pili to the bacterial cell wall. FACS analysis of the ΔSrtC1 strain SF370 confirmed that deletion of SrtC1 does not eliminate cell surface expression of M1_126, M1_128 or M1_130. See Figure 179 G-I, which show a shift in fluorescence when antibodies immunoreactive to M1_126 (Figure 179 G), M1_128 (Figure 179 H), and M1_130 (Figure 179 I) are used to detect cell surface protein expression on ΔSrtC1 bacteria. Thus, SrtC1 deletion prevents pilus formation, but not surface anchoring of proteins involved in pilus formation on the surface of bacteria. Another sortase is possibly involved in anchoring of the proteins to the bacteria surface. Pilus polymerization in *C. diphtheriae* is also dependent on particular sortase enzyme whose gene resides at the same genetic locus as the pilus components (*7, 8*).

The LepA signal peptidase, Spy0127, also appears to be essential for pilus assembly in strain SF370. LepA deletion mutants (ΔLepA) of strain SF370 fail to assemble pili on the cell surface. Not only are the ΔLepA mutants unable to assemble pili, they are also deficient at cell surface M1 expression. See Figure 180, which provides a FACS analysis of the wildtype (A) and ΔLepA mutant (B) SF370 bacteria using M1 antisera. No shift in fluorescence is observed for the ΔLepA mutant bacteria in the presence of M1 immune serum. It is possible that these deletion mutants of LepA will be useful for detecting non-M, non-pili, surface exposed antigens on the surface of GAS, or any Gram positive bacteria. These antigens may also be useful in immunogenic compositions.

Pili were also observed in M5 strain ISS4882 and M12 strain 20010296. The M5 strain ISS4882 contains genes for four predicted surface exposed proteins (GAS AI-3). Antisera against three of the four products of the FCT region (GAS AI-3) of M5_ISS4883 (Cpa, M5_orf80, M5_orf82) stained high molecular weight ladders in Western blot analysis (Figure 163 C). Long pili were visible when antisera against M5_orf80 was used in immunogold staining followed by electron microscopy (Figure 163L).

The M12 strain 20010296 contains genes for five predicted surface exposed proteins. (GAS AI-4) Antisera against three of the five products of the FCT region (GAS AI-4) of M12_20010296 (Cpa, EftLSL.A, Orf2) stained high molecular weight ladders in Westen blot analysis (Figure 163 D). Long pili were visible when antisera against EftLSL.A were used (Fig. 163 M).

The major pilus forming proteins identified in the four strains studied by applicants (T6, M1_128, M5_orf80 and EftLSL.A) share between 23% and 65% amino acid identity in any pairwise comparison, indicating that each pilus may represent a different Lancefield T-antigen. Each pilus is part of a trypsin resistant structure on the GAS bacteria surface, as is the case for the Lancefield T-antigens. See Figure 165, which provides a FACS analysis of bacteria harboring each of the FCT types that had or had not been treated with trypsin (*6*). Following treatment, surface expression of the pilus proteins was assayed by indirect immunofluorescence and flow cytometry using antibodies specific for the pilus proteins, the bacteria's respective M proteins, or surface proteins not associated with the pili (Figure 165). Staining the cells with sera specific for proteins associated with the pili was not effected by trypsin treatment, whereas trypsin treatment substantially reduced detection of M-proteins or surface proteins not associated with pili.

The pili structures identified on the surface of the GAS bacteria were confirmed to be Lancefield T antigens when commercially available T-serotyping sera detected the pili on the surface of bacteria. Western blot analysis was initially performed to determine if polyvalent serum pools (designated T, U, W, X, andY) could detect recombinant proteins for each of the major pilis components (T6, M1_128, M5_orf80 and EftLSL.A) identified in the strains of bacteria discussed above. Pool U, which contains the T6 serum, recognized the T6 protein specifically (a surface exposed pilus protein from GAS AI-1)(Fig. 166 B). Pool T specifically recognized M1_128 (a surface exposed pilus protein from GAS AI-2) (Fig. 166 A). Pool W recognized both M5_orf80 and EftLSL.A (Fig. 166 C). Using monovalent sera representative of each of the components of each polyvalent pool, applicants confirmed the specificity of the T6 antigen (corresponding to a surface exposed pilus protein from GAS AI-1)(Fig. 166 E) and identified M1_128 as antigen T1 (corresponding to a surface exposed pilus protein from GAS AI-2) (Fig. 166 D), EftLSL.A as antigen T12 (corresponding to a surface exposed pilus protein from GAS AI-4) (Fig. 166 G) and M5_orf80 as a common antigen recognized by the related sera T5, T27 and T44 (corresponding to a surface exposed pilus protein from GAS AI-3).

Confirming applicants observations, discussed above, that deleting the M1_128 gene from M1_SF370 abolishes pilus formation, the pool T sera stained whole M1_SF370 bacteria (Fig. 166 H) but failed to stain M1_SF370 bacteria lacking the M1_128 gene (Fig. 166 I).

As discussed above, Applicants have identified at least four different Group A Streptococcus Adhesin Islands. While these GAS AI sequences can be identified in numerous M types, Applicants have surprisingly discovered a correlation between the four main pilus subunits from the four different GAS AI types and specific T classifications. While other trypsin-resistant surface exposed proteins are likely also implicated in the T classification designations, the discovery of the role of the GAS adhesin islands (and the associated hyper-oligomeric pilus like structures) in T classification and GAS serotype variance has important implications for prevention and treatment of GAS infections. Applicants have identified protein components within each of the GAS adhesin islands which are associated with the pilus formation. These proteins are believed to be involved in the bacteria's initial adherence mechanisms. Immunological recognition of these proteins may allow the host immune response to slow or prevent the bacteria's transition into the more pathogenic later stages of infection. In addition, the GAS pili may be involved in formation of biofilms. Applicants have discovered that the GBS pili structures appear to be implicated in the formation of biofilms (populations of bacteria growing on a surface, often enclosed in an exopolysaccharide matrix). Biofilms are generally associated with bacterial resistance, as antibiotic treatments and host immune response are frequently unable to erradicate all of the bacteria components of the biofilm. Direction of a host immune response against surface proteins exposed during the first steps of bacterial attachment (i.e., before complete biofilm formation) is preferable.

The invention therefore provides for improved immunogenic compositions against GAS infection which may target GAS bacteria during their initial attachment efforts to the host epithelial cells and may provide protection against a wide range of GAS serotypes. The immunogenic compositions of the invention include GAS AI surface proteins which may be formulated in an oligomeric, or hyperoligomeric (pilus) form. The invention also includes combinations of GAS AI surface proteins. Combinations of GAS AI surface proteins may be selected from the same adhesin island or they may be selected from different GAS adhesin islands.

The invention comprises compositions comprising a first GAS AI protein and a second GAS AI protein wherein the first and second GAS AI proteins are derived from different GAS adhesin islands. For example, the invention includes a composition comprising at least two GAS AI proteins wherein the GAS AI proteins are encoded by the adhesin islands selected from the group consisting of GAS AI-1 and AI-2; GAS AI-1 and GAS AI-3; GAS AI-1 and GAS AI-4; GAS AI-2 and GAS AI-3; GAS AI-2 and GAS AI-4; and GAS AI-3 and GAS AI-4. Preferably the two GAS AI proteins are derived from different T-types.

A schematic arrangement of GAS Adhesin Island sequences is set forth in FIGURE 162. In all strains, the AI region is flanked by the highly conserved open reading frames M1_123 and M1-136. Between three and five genes in each locus code for surface proteins containing LPXTG motifs. These surface proteins also all belong to the family of genes coding for ECM binding adhesins.

Adhesin island sequences can be identified in numerous M types of Group A Streptococcus. Examples of AI sequences within M1, M6, M3, M5, M12, M18, and M49 serotypes are discussed below.

GAS Adhesin Islands generally include a series of open reading frames within a GAS genome that encode for a collection of surface proteins and sortases. A GAS Adhesin Island may encode for amino acid sequences comprising at least one surface protein. Alternatively, a GAS Adhesin Island may encode for at least two surface proteins and at least one sortase. Preferably, a GAS Adhesin Island encodes for at least three surface proteins and at least two sortases. One or more of the surface proteins may include an LPXTG motif (such as LPXTG (SEQ ID NO: 122)) or other sortase substrate motif. One or more GAS AI surface proteins may participate in the formation of a pilus structure on the surface of the Gram positive bacteria.

GAS Adhesin Islands of the invention preferably include a divergently transcribed transcriptional regulator. The transcriptional regulator may regulate the expression of the GAS AI operon. Examples of transcriptional regulators found in GAS AI sequences include *RofA* and *Nra*.

The GAS AI surface proteins may bind or otherwise adhere to fibrinogen, fibronectin, or collagen. One or more of the GAS AI surface proteins may comprise a fimbrial structural subunit.

One or more of the GAS AI surface proteins may include an LPXTG motif or other sortase substrate motif. The LPXTG motif may be followed by a hydrophobic region and a charged C terminus, which are thought to retard the protein in the cell membrane to facilitate recognition by the membrane-localized sortase. See Barnett, et al., J. Bacteriology (2004) 186 (17): 5865-5875.

GAS AI sequences may be generally categorized as Type 1, Type 2, Type 3, or Type 4, depending on the number and type of sortase sequences within the island and the percentage identity of other proteins (with the exception of RofA and cpa) within the island. Figure 167 provides a chart indicating the number and type of sortase sequences identified within the adhesin islands of various strains and serotypes of GAS. As can be seen in this figure, all GAS strains and serotypes thus far characterized as an AI-1 have a SrtB type sortase, all GAS strains and serotypes thus far characterized as an AI-2 have SrtB and SrtC1 type sortases, all GAS strains and serotypes thus far characterized as an AI-3 have a SrtC2 type sortase, and all GAS strains and serotypes thus far characterized as an AI-4 have SrtB and SrtC2 type sortases. A comparison of the percentage identity of sequences within the adhesin islands was presented in Table 45, see above.

(1) <u>Adhesin Island sequence within M6: GAS Adhesin Island 1 ("GAS AI-1")</u>

A GAS Adhesin Island within M6 serotype (MGAS10394) is outlined in Table 4 below. This GAS adhesin island 1 ("GAS AI-1") comprises surface proteins, a srtB sortase and a rofA divergently transcribed transcriptional regulator.

GAS AI-1 surface proteins include Spy0157 (a fibronectin binding protein), Spy0159 (a collagen adhesion protein) and Spy0160 (a fimbrial structural subunit). Preferably, each of these GAS AI-1 surface proteins includes an LPXTG sortase substrate motif, such as LPXTG (SEQ ID NO: 122) or LPXSG (SEQ ID NO: 134) (conservative replacement of threonine with serine).

GAS AI-1 includes a srtB type sortase. GAS srtB sortases may preferably anchor surface proteins with an LPSTG motif (SEQ ID NO: 166), particularly where the motif is followed by a serine.

Table 4: GAS AI-1 sequences from M6 isolate (MGAS10394)

AI-1 sequence identifier	Sortase substrate sequence or sortase type	functional description
M6_Spy0156		Transcriptional regulator (*rofA*)
M6_Spy0157	LPXTG	Fibronectin-binding protein
M6_Spy0158		Reverse transcriptase
M6_Spy0159	LPXSG	Collagen adhesion protein
M6_Spy0160	LPXTG	Fimbrial structural subunit
M6_Spy0161	srtB	Sortase

M6_Spy0160 appears to be present on the surface of GAS as part of oligomeric (pilus) structures. Figures 127-132 present electron micrographs of GAS serotype M6 strain 3650 immunogold stained for M6_Spy0160 using anti-M6_Spy0160 antiserum. Oligomeric or hyperoligomeric structures labelled with gold particles can be seen extending from the surface of the GAS in each of these figures, indicating the presence of multiple M6_Spy0160 polypeptides in the oligomeric or hyperoligomeric structures. Figure 176 A-F present electron micrographs of GAS M6 strain 2724 immunogold stained for M6_Spy0160 using anti-M6_Spy0160 antiserum (Figures 176 A-E) or immunogold stained for M6_Spy0159 using anti-M6_Spy0159 antiserum (Figure 176 F). Oligomeric or hyperoligomeric structures labelled with gold particles can again be seen extending from the surface of the M6 strain 2724 GAS bacteria immunogold stained for M6_Spy0160. M6_Spy0159 is also detected on the surface of the M6 strain 2724 GAS.

FACS analysis has confirmed that the GAS AI-1 surface proteins spyM6_0159 and spyM6_0160 are indeed expressed on the surface of GAS. Figure 73 provides the results of FACS analysis for surface expression of spyM6_0159 on each of GAS serotypes M6 2724, M6 3650, and M6 2894. A shift in fluorescence is observed for each GAS serotype when anti-spyM6_0159 antiserum is present, demonstrating cell surface expression. Table 18, below, quantitatively summarizes the FACS fluorescence values obtained for each GAS serotype in the presence of pre-immune antiserum, anti-spyM6_0159 antiserum, and the difference in fluorescence value between the pre-immune and anti-spyM6_0159 antiserum.

Table 18: Summary of FACS values for surface expression of spyM6_0159

2724			3650			2894		
Pre-immune	Anti-spyM6_0159	Change	Pre-immune	Anti-spyM6_0159	Change	Pre-immune	Anti-spyM6_0159	Change
134.84	427.48	293	149.68	712.62	563	193.86	597.8	404

Figure 74 provides the results of FACS analysis for surface expression of spyM6_0160 on each of GAS serotypes M6 2724, M6 3650, and M6 2894. In the presence of of anti-spyM6_0160 antiserum, a shift in fluorescence is observed for each GAS serotype, which demonstrates its cell surface expression. Table 19, below, quantitatively summarizes the FACS fluorescence values obtained for each GAS serotype in the presence of pre-immune antiserum, anti-spyM6_0160 antiserum, and the change in fluorescence value between the pre-immune and anti-spyM6_0160 antiserum.

Table 19: Summary of FACS values for surface expression of spyM6_0160

2724			3650			2894		
Pre-immune	Anti-spyM6_0160	change	Pre-immune	Anti-spyM6_0160	change	Pre-immune	Anti-spyM6_0160	change
117.12	443.24	326	128.57	776.39	648	125.87	621.17	495

Surface expression of M6_Spy0159 and M6_Spy0160 on M6 serotype GAS has also been confirmed by Western blot analysis. Figure 98 shows that while pre-immune sera (P α-0159) does not detect expression of M6_Spy0159 in GAS serotype M6, anti-M6_Spy0159 immune sera (I α-0159) is able to detect M6_Spy0159 protein in both total GAS M6 extracts (M6 tot) and GAS M6 fractions enriched for cell surface proteins (M6 surf prot). The M6_Spy0159 proteins detected in the total GAS M6 extracts or the GAS M6 extracts enriched for surface proteins are also present as high molecular weight structures, indicating that M6_Spy0159 may be in an oligomeric (pilus) form.

Figure 112 shows that while preimmune sera (Preimmune Anti 106) does not detect expression of M6_Spy0160 in GAS serotype M6 strain 2724, anti-M6_Spy0160 immune sera (Anti 160) does in both total GAS M6 strain 2724 extracts (M6 2724 tot) and GAS M6 strain 2724 fractions enriched for surface proteins. The M6_Spy0160 proteins detected in the total GAS M6 strain 2724 extracts or the GAS M6 strain 2724 extracts enriched for surface proteins are also present as high molecular weight structures, indicating that M6_Spy0160 may be in an oligomeric (pilus) form.

Figures 110 and 111 both further verify the presence of M6_Spy0159 and M6_Spy0160 in higher molecular weight structures on the surface of GAS. Figure 110 provides a Western blot performed to detect M6_Spy0159 and M6_Spy0160 in GAS M6 strain 2724 extracts enriched for surface proteins. Antiserum raised against either M6_Spy0159 (Anti-159) or M6_Spy0160 (Anti-160) cross-hybridizes with high molecular weight structures (pili) in these extracts. Figure 111 provides a similar Western blot that verifies the presence of M6_ Spy0159 and M6_Spy0160 in high molecular weight structures in GAS M6 strain 3650 extracts enriched for surface proteins.

SpyM6_0157 (a fibronectin-binding protein) may also be expressed on the surface of GAS serotype M6 bacteria. Figure 174 shows the results of FACS analysis for surface expression of spyM6_0157 on M6 strain 3650. A slight shift in fluorescence is observed, which demonstrates that some spyM6_0157 may be expressed on the GAS cell surface.

Adhesin Island sequence within M6: GAS Adhesin Island 2 (“GAS AI-2”)

A GAS Adhesin Island within M1 serotype (SF370) is outlined in Table 5 below. This GAS adhesin island 2 (“GAS AI-2”) comprises surface proteins, a SrtB sortase, a SrtC1 sortase and a RofA divergently transcribed transcriptional regulator.

GAS AI-2 surface proteins include GAS 15 (Cpa), Spy0128 (thought to be a fimbrial protein) and Spy0130 (a hypothetical protein). Preferably, each of these GAS AI-2 surface proteins includes an LPXTG sortase substrate motif, such as LPXTG (SEQ ID NO: 122), VVXTG (SEQ ID NO: 135), or EVXTG (SEQ ID NO: 136).

GAS AI-2 includes a srtB type sortase and a srtC1 sortase. As discussed above, GAS SrtB sortases may preferably anchor surface proteins with an LPSTG (SEQ ID NO: 166) motif, particularly where the motif is followed by a serine. GAS SrtC1 sortase may preferentially anchor surface proteins with a V(P/V)PTG (SEQ ID NO: 167) motif. GAS SrtC1 may be differentially regulated by RofA.

GAS AI-2 may also include a LepA putative signal peptidase I protein.

Table 5 : GAS AI-2 sequence from M1 isolate (SF370)

AI-2 sequence identifier	Sortase substrate sequence or sortase type	functional description
SPy0124		rofA regulatory protein
GAS15(not annotated in SF370)	VVXTG	cpa

SPy0127		LepA putative signal peptidase I
SPy0128 (GAS16)	EVXTG	hypothetical protein (fimbrial)
SPy0129 (GAS17)	srtC1	sortase
SPy0130 (GAS18)	LPXTG	hypothetical protein
SPy0131		conserved hypothetical protein
SPy0133		conserved hypothetical protein
SPy0135 (GAS20)	srtB	sortase (putative fimbrial-associated protein)

GAS 15, GAS 16, and GAS 18 appear to be present on the surface of GAS as part of oligomeric (pilus) structures. Figures 113-115 present electron micrographs of GAS serotype M1 strain SF370 immunogold stained for GAS 15 using anti-GAS 15 antiserum. Figures 116-121 provide electron micrographs of GAS serotype M1 strain SF370 immunogold stained for GAS 16 using anti-GAS 16 antiserum. Figures 122-125 present electron micrograph of GAS serotype M1 strain SF370 immunogold stained for GAS 18 using anti-GAS 18 antiserum. Oligomers of these proteins can be seen on the surface of SF370 bacteria in the immuno-gold stained micrographs.

Figure 126 reveals a hyperoligomer on the surface of a GAS serotype M1 strain SF370 bacterium immunogold stained for GAS 18. This long hyperoliogmeric structure comprising GAS 18 stretches far out into the supernatant from the surface of the bacteria.

FACS analysis has confirmed that the GAS AI-2 surface proteins GAS 15, GAS 16, and GAS 18 are expressed on the surface of GAS. Figure 75 provides the results of FACS analysis for surface expression of GAS 15 on each of GAS serotypes M1 2719, M1 2580, M1 3280, M1 SF370, M1 2913, and M1 3348. A shift in fluorescence is observed for each GAS serotype when anti-GAS 15 antiserum is present, demonstrating cell surface expression. Table 20, below, quantitatively summarizes the FACS fluorescence values obtained for each GAS serotype in the presence of pre-immune antiserum, anti-GAS 15 antiserum, and the difference in fluorescence value between the pre-immune and anti-GAS 15 antiserum.

Table 20: Summary of FACS values for surface expression of GAS 15

2719			2580			3280		
Pre-immune	Anti-GAS 15	Change	Pre-immune	Anti-GAS 15	Change	Pre-immune	Anti-GAS 15	Change
159.46	712.71	553	123.9	682.84	559	217.02	639.69	423

SF370			2913			3348		
Pre-immune	Anti-GAS 15	Change	Pre-immune	Anti-GAS 15	Change	Pre-immune	Anti-GAS 15	Change
201.93	722.68	521	121.41	600.45	479	152.09	446.41	294

Figures 76 and 79 provide the results of FACS analysis for surface expression of GAS 16 on each of GAS serotypes M1 2719, M1 2580, M1 3280, M1 SF370, M1 2913, and M1 3348. The FACS data in Figure 76 was obtained using antisera was raised against full length GAS 16. In the presence of this anti-GAS 16 antiserum, a shift in fluorescence is observed for each GAS serotype, demonstrating its cell surface expression. Table 21, below, quantitatively summarizes the FACS fluorescence values obtained for each GAS serotype in the presence of pre-immune antiserum, anti-GAS 16 antiserum, and the change in fluorescence value between the pre-immune and anti-GAS 16 antiserum.

Table 21: Summary of FACS values for surface expression of GAS 16

2719			2580			3280		
Pre-immune	Anti-GAS 16	Change	Pre-immune	Anti-GAS 16	Change	Pre-immune	Anti-GAS 16	Change
233.27	690.09	457	133.82	732.29	598	264.47	649.43	385
SF370			2913			3348		
Pre-immune	Anti-GAS 16	Change	Pre-immune	Anti-GAS 16	Change	Pre-immune	Anti-GAS 16	Change
237.2	727.46	490	138.52	588.04	450	180.56	420.93	240

The FACS data in Figure 79 was obtained using antisera was raised against a truncated GAS 16, which is encoded by SEQ ID NO: 179, shown below.

SEQ ID NO: 179:

```
GCTACAACAGTTCACGGGGAGACTGTTGTAAACGGAGCCAAACTAACAGTTACAAAAAACCTTGATTTAGTTAAT
AGCAATGCATTAATTCCAAATACAGATTTTACATTTAAAATCGAACCTGATACTACTGTCAACGAAGACGGAAAT
AAGTTTAAAGGTGTAGCTTTGAACACACCGATGACTAAAGTCACTTACACCAATTCAGATAAAGGTGGATCAAAT
ACGAAAACTGCAGAATTTGATTTTTCAGAAGTTACTTTTGAAAAACCAGGTGTTTATTATTACAAAGTAACTGAG
GAGAAGATAGATAAAGTTCCTGGTGTTTCTTATGATACAACATCTTACACTGTTCAAGTTCATGTCTTGTGGAAT
GAAGAGCAACAAAAACCAGTAGCTACTTATATTGTTGGTTATAAAGAAGGTAGTAAGGTGCCAATTCAGTTCAAA
AATAGCTTAGATTCTACTACATTAACGGTGAAGAAAAAAGTTTCAGGTACCGGTGGAGATCGCTCTAAAGATTTT
AATTTTGGTCTGACTTTAAAAGCAAATCAGTATTATAAGGCGTCAGAAAAAGTCATGATTGAGAAGACAACTAAA
GGTGGTCAAGCTCCTGTTCAAACAGAGGCTAGTATAGATCAACTCTATCATTTTACCTTGAAAGATGGTGAATCA
ATCAAAGTCACAAATCTTCCAGTAGGTGTGGATTATGTTGTCACTGAAGACGATTACAAATCAGAAAAATATACA
ACCAACGTGGAAGTTAGTCCTCAAGATGGAGCTGTAAAAAATATCGCAGGTAATTCAACTGAACAAGAGACATCT
ACTGATAAAGATATGACCATTACTTTTACAAATAAAAAAGATTT
```

In the presence of this anti-GAS 16 antiserum, a shift in fluorescence is observed for each GAS serotype, demonstrating its cell surface expression. Table 22, below, quantitatively summarizes the FACS fluorescence values obtained for each GAS serotype in the presence of pre-immune antiserum, anti-GAS 16 antiserum, and the change in fluorescence value between the pre-immune and anti-GAS 16 antiserum.

Table 22: Summary of FACS values for surface expression of GAS 16 using a second antisera

2719			2580			3280		
Pre-immune	Anti-GAS 16	Change	Pre-immune	Anti-GAS 16	Change	Pre-immune	Anti-GAS 16	Change
141.55	650.22	509	119.57	672.35	553	209.18	666.71	458
SF370			2913			3348		
Pre-immune	Anti-GAS 16	Change	Pre-immune	Anti-GAS 16	Change	Pre-immune	Anti-GAS 16	Change
159.92	719.32	559	115.97	585.9	470	146.1	414.01	268

Figures 77 and 78 provide the results of FACS analysis for surface expression of GAS 18 on each of GAS serotypes M1 2719, M1 2580, M1 3280, M1 SF370, M1 2913, and M1 3348. The antiserum used to obtain the FACS data in each of Figures 77 and 78 was different, although each was raised against full length GAS 18. In the presence of each of the anti-GAS 18 antisera, a shift in fluorescence is observed for each GAS serotype, demonstrating its cell surface expression. Tables 23 and 24, below, quantitatively summarizes the FACS fluorescence values obtained for each GAS serotype in the presence of pre-immune antiserum, first or second anti-GAS 18 antiserum, and the change in fluorescence value between the pre-immune and first or second anti-GAS 18 antiserum.

Table 23: Summary of FACS values for surface expression of GAS 18

2719			2580			3280		
Pre-immune	Anti-GAS 18	Change	Pre-immune	Anti-GAS 18	Change	Pre-immune	Anti-GAS 18	Change
135.68	327.98	192	116.32	379.41	263	208.12	380.84	173

SF370			2913			3348		
Pre-immune	Anti-GAS 18	Change	Pre-immune	Anti-GAS 18	Change	Pre-immune	Anti-GAS 18	Change
185.39	438.23	253	119.95	373.32	253	147.12	266.51	119

Table 24: Summary of FACS values for surface expression of GAS 18 using a second antisera

2719			2580			3280		
Pre-immune	Anti-GAS 18	Change	Pre-immune	Anti-GAS 18	Change	Pre-immune	Anti-GAS 18	Change
150.4	250.39	100	139.18	386.38	247	253.38	347.72	94

SF370			2913			3348		
Pre-immune	Anti-GAS 18	Change	Pre-immune	Anti-GAS 18	Change	Pre-immune	Anti-GAS 18	Change
188.64	373.11	184	124.94	384.82	260	168.8	213.65	45

Surface expression of GAS 15, GAS 16, and GAS 18 on M1 serotype GAS has also been confirmed by Western blot analysis. Figure 89 shows that while pre-immune sera does not detect GAS M1 expression of GAS 15, anti-GAS 15 immune sera is able to detect GAS 15 protein in both total GAS M1 extracts and GAS M1 proteins enriched for cell surface proteins. The GAS 15 proteins detected in the M1 extracts enriched for surface proteins are also present as high molecular weight structures, indicating that GAS 15 may be in an oligomeric (pilus) form. Figure 90 also shows the results of Western blot analysis of M1 serotype GAS using anti-GAS 15 antisera. Again, the lanes that contain GAS M1 extracts enriched for surface proteins (M1 prot sup) show the presence of high molecular weight structures that may be oligomers of GAS 15. Figure 91 provides an additional Western blot identical to that of Figure 90, but that was probed with pre-immune sera. As expected, no proteins were detected on this membrane.

Figure 92 provides a Western blot that was probed for GAS 16 protein. While pre-immune sera does not detect GAS M1 expression of GAS 16, anti-GAS 16 immune sera is able to detect GAS

16 protein in GAS M1 extracts enriched for cell surface proteins. The GAS 16 proteins detected in the M1 extracts enriched for surface proteins are present as high molecular weight structures, indicating that GAS 16 may be in an oligomeric (pilus) form. Figure 93 also shows the results of Western blot analysis of M1 serotype GAS using anti-GAS 16 antisera. The lanes that contain total GAS M1 protein (M1 tot new and M1 tot old) and the lane that contains GAS M1 extracts enriched for surface proteins (M1 prot sup) show the presence of high molecular weight structures that may be oligomers of GAS 16. Figure 94 provides an additional Western blot identical to that of Figure 93, but that was probed with pre-immune sera. As expected, no proteins were detected on this membrane.

Figure 95 provides a Western blot that was probed for GAS 18 protein. While pre-immune sera does not detect GAS M1 expression of GAS 18, anti-GAS 18 immune sera is able to detect GAS 18 protein in GAS M1 extracts enriched for cell surface proteins. The GAS 18 proteins detected in the M1 extracts enriched for surface proteins are present as high molecular weight structures, indicating that GAS 18 may be in an oligomeric (pilus) form. Figure 96 also shows the results of Western blot analysis of M1 serotype GAS using anti-GAS 18 antisera. The lane that contains GAS M1 extracts enriched for surface proteins (M1 prot sup) show the presence of high molecular weight structures that may be oligomers of GAS 18. Figure 97 provides an additional Western blot identical to that of Figure 96, but that was probed with pre-immune sera. As expected, no proteins were detected on this membrane.

Figures 102-106 provide additional Western blots to verify the presence of GAS 15, GAS 16, and GAS 18 in high molecular weight structures in GAS. Each Western blot was performed using proteins from a different GAS M1 strain, 2580, 2913, 3280, 3348, and 2719. Each Western blot was probed with antisera raised against each of GAS 15, GAS 16, and GAS 18. As can be seen in Figures 102-106, none of the Western blots shows detection of proteins using pre-immune serum (Pα-158, Pα-15, Pα-16, or Pα-18), while each Western blot shows cross-hybridization of the GAS 15 (Iα-15), GAS 16 (Iα-16), and GAS 18 (Iα-18) antisera to high molecular weight structures. Thus, these Western blots confirm that GAS 15, GAS 16, and GAS 18 can be present in pili in GAS M1.

Figure 107 provides a similar Western blot performed to detect GAS 15, GAS 16, and GAS 18 proteins in a GAS serotype M1 strain SF370 protein fraction enriched for surface proteins. This Western blot also shows detection of GAS 15 (Anti-15), GAS 16 (Anti-16), and GAS 18 (Anti-18) as high molecular weight structures.

(3) Adhesin Island sequence within M3, M5, and M18: GAS Adhesin Island 3 ("GAS AI-3")

GAS Adhesin Island sequences within M3, M5, and M18 serotypes are outlined in Tables 6 – 8 and 10 below. This GAS adhesin island 3 ("GAS AI-3") comprises surface proteins, a SrtC2 sortase, and a Negative transcriptional regulator (Nra) divergently transcribed transcriptional regulator.

GAS AI-3 surface proteins within include a collagen binding protein, a fimbrial protein, a F2 like fibronectin-binding protein. GAS AI-3 surface proteins may also include a hypothetical surface protein. Preferably, each of these GAS AI-3 surface proteins include an LPXTG sortase substrate motif, such as LPXTG (SEQ ID NO: 122), VPXTG (SEQ ID NO: 137), QVXTG (SEQ ID NO: 138) or LPXAG (SEQ ID NO: 139).

GAS AI-3 includes a SrtC2 type sortase. GAS SrtC2 type sortases may preferably anchor surface proteins with a QVPTG (SEQ ID NO: 140) motif, particularly when the motif is followed by a hydrophobic region and a charged C terminus tail. GAS SrtC2 may be differentially regulated by Nra.

GAS AI-3 may also include a LepA putative signal peptidase I protein.

GAS AI-3 may also include a putative multiple sugar metabolism regulator.

Table 6: GAS AI-3 sequences from M3 isolate (MGAS315)

AI-3 sequence identifier	Sortase substrate sequence or sortase type	Functional description
SpyM3_0097		Negative transcriptional regulator (Nra)
SpyM3_0098	VPXTG	putative collagen binding protein (Cpb)
SpyM3_0099		LepA putative signal peptidase I
SpyM3_0100	QVXTG	conserved hypothetical protein (fimbrial)
SpyM3_0101	SrtC2	sortase
SpyM3_0102	LPXAG	hypothetical protein
SpyM3_0103		putative multiple sugar metabolism regulator
SpyM3_0104	LPXTG	protein F2 like fibronectin-binding protein

Table 7: GAS AI-3 sequence from M3 isolate (SSI-1)

AI-3 sequence identifier	Sortase Substrate sequence or sortase type	Functional description
SPs0099		Negative transcriptional regulator (Nra)
SPs0100	VPXTG	putative collagen binding protein (Cpb)
SPs0101		LepA putative signal peptidase I
SPs0102	QVXTG	conserved hypothetical protein (fimbrial)
SPs0103	SrtC2	sortase
SPs0104	LPXAG	hypothetical protein
SPs0105		putative multiple sugar metabolism regulator
SPs0106	LPXTG	protein F2 like fibronectin-binding protein

Table 10: GAS AI-3 sequences from M5 isolate (Manfredo)

AI-3 sequence identifier	Sortase substrate sequence or sortase type	Functional description
orf77		Negative transcriptional regulator (Nra)
orf78	VPXTG	putative collagen binding protein (Cpb)
orf79		LepA putative signal peptidase I
orf80	QVXTG	conserved hypothetical protein (fimbrial)
orf81	SrtC2	sortase

orf82	LPXAG	hypothetical protein
orf83		putative multiple sugar metabolism regulator
orf84	LPXTG	protein F2 like fibronectin-binding protein

Table 8: GAS AI-3 sequences from M18 isolate (MGAS8232)

AI-3 sequence identifier	Sortase substrate sequence or sortase type	Functional description
spyM18_0125		Negative transcriptional regulator (Nra) (N-terminal fragment)
spyM18_0126	VPXTG	putative collagen binding protein (Cpb)
spyM18_0127		LepA putative signal peptidase I
spyM18_0128	QVXTG	conserved hypothetical protein (fimbrial)
spyM18_0129	SrtC2	sortase
spyM18_0130	LPXAG	hypothetical protein
spyM18_0131		putative multiple sugar metabolism regulator
spyM18_0132	LPXTG	protein F2 like fibronectin-binding protein

Table 44: GAS AI-3 sequences from M49 isolate (591)

AI-3 sequence identifier	Sortase substrate sequence or sortase type	Functional description
SpyoM01000156		Negative transcriptional regulator (Nra)
SpyoM01000155	VPXTG	collagen binding protein (Cpa)
SpyoM01000154		LepA putative signal peptidase I
SpyoM01000153	QVXTG	conserved hypothetical protein (fimbrial)
SpyoM01000152	SrtC2	sortase
SpyoM01000151	LPXAG	hypothetical protein
SpyoM01000150		MsmRL
SpyoM01000149	LPXTG	protein F2 like fibronectin-binding protein

A schematic of AI-3 serotypes M3, M5, M18, and M49 is shown in Figure 51A. Each contains an open reading frame encoding a SrtC2-type sortase of nearly identical amino acid sequence. See Figure 52B for an amino acid sequence alignment for each of the SrtC2 amino acid sequences.

The protein F2-like fibronectin-binding protein of each these type 3 adhesin islands contains a pilin motif and an E-box. Figure 60 indicates the amino acid sequence of the pilin motif and E-box of each of GAS AI-3 serotype M3 MGAS315 (SpyM3_0104/21909640), GAS AI-3 serotype M3 SSI (Sps0106/28895018), GAS AI-3 serotype M18 (SpyM18_0132/19745307), and GASAI-3 serotype M5 (orf84).

FACS analysis has confirmed that the GAS AI-3 surface proteins SpyM3_0098, SpyM3_0100, SpyM3_0102, and SpyM3_0104 are expressed on the surface of GAS. Figure 80 provides the results of FACS analysis for surface expression of SpyM3_0098 on each of GAS serotypes M3 2721 and M3 3135. A shift in fluorescence is observed for each GAS serotype when anti-SpyM3_0098 antiserum is present, demonstrating cell surface expression. Table 25, below, quantitatively summarizes the FACS fluorescence values obtained for each GAS serotype in the presence of pre-immune antiserum, anti-SpyM3_0098 antiserum, and the difference in fluorescence value between the pre-immune and anti-SpyM3_0098 antiserum.

Table 25: Summary of FACS values for surface expression of SpyM3_0098

2721			3135		
Pre-immune	Anti-spyM3_0098	Change	Pre-immune	Anti-spyM3_0098	Change
117.85	249.51	132	99.17	277.21	178

Figure 81 provides the results of FACS analysis for surface expression of SpyM3_0100 on each of GAS serotypes M3 2721 and M3 3135. A shift in fluorescence is observed for each GAS serotype when anti-SpyM3_0100 antiserum is present, demonstrating cell surface expression. Table 26, below, quantitatively summarizes the FACS fluorescence values obtained for each GAS serotype in the presence of pre-immune antiserum, anti-SpyM3_0100 antiserum, and the difference in fluorescence value between the pre-immune and anti-SpyM3_0100 antiserum.

Table 26: Summary of FACS values for surface expression of SpyM3_0100

2721			3135		
Pre-immune	Anti-spyM3_0100	Change	Pre-immune	Anti-spyM3_0100	Change
110.31	181.91	72	97.87	250.01	152

Figure 82 provides the results of FACS analysis for surface expression of SpyM3_0102 on each of GAS serotypes M3 2721 and M3 3135. A shift in fluorescence is observed for each GAS serotype when anti-SpyM3_0102 antiserum is present, demonstrating cell surface expression. Table 27, below, quantitatively summarizes the FACS fluorescence values obtained for each GAS serotype in the presence of pre-immune antiserum, anti-SpyM3_0102 antiserum, and the difference in fluorescence value between the pre-immune and anti-SpyM3_0102 antiserum.

Table 27: Summary of FACS values for surface expression of SpyM3_0102 in M3 serotypes

2721			3135		
Pre-immune	Anti-spyM3_0102	Change	Pre-immune	Anti-spyM3_0102	Change
109.86	155.26	45	100.02	112.58	13

Figure 82 also provides the results of FACS analysis for surface expression of a pilin antigen that has homology to SpyM3_0102 identified in a different GAS serotype, M6. FACS analysis conducted with the SpyM3_0102 antisera was able to detect surface expression of the homologous SpyM3_0102 antigen on each of GAS serotypes M6 2724, M6 3650, and M6 2894. Table 28, below, quantitatively summarizes the FACS fluorescence values obtained for each GAS serotype in the presence of pre-immune antiserum, anti-SpyM3_0102 antiserum, and the difference in fluorescence value between the pre-immune and anti-SpyM3_0102 antiserum.

Table 28: Summary of FACS values for surface expression of SpyM3_0102 in M6 serotypes

2724	3650	2894

Pre-immune	Anti-spyM3_0102	Change	Pre-immune	Anti-spyM3_0102	Change	Pre-immune	Anti-spyM3_0102	Change
146.59	254.03	107	162.56	294.03	131	175.49	313.69	138

SpyM3_0102 is also homologous to pilin antigen 19224139 of GAS serotype M12. Antisera raised against SpyM3_0102 is able to detect high molecular weight structures in GAS serotype M12 strain 2728 protein fractions enriched for surface proteins, which would contain the 19224139 antigen. See Figure 109 at the lane labelled M12 2728 surf prot.

Figure 83 provides the results of FACS analysis for surface expression of SpyM3_0104 on each of GAS serotypes M3 2721 and M3 3135. A shift in fluorescence is observed for each GAS serotype when anti-SpyM3_0104 antiserum is present, demonstrating cell surface expression. Table 29, below, quantitatively summarizes the FACS fluorescence values obtained for each GAS serotype in the presence of pre-immune antiserum, anti-SpyM3_0104 antiserum, and the difference in fluorescence value between the pre-immune and anti-SpyM3_0104 antiserum.

Table 29: Summary of FACS values for surface expression of SpyM3_0104 in M3 serotypes

2721			3135		
Pre-immune	Anti-spyM3_0104	Change	Pre-immune	Anti-spyM3_0104	Change
128.45	351.65	223	105.1	339.88	235

Figure 83 also provides the results of FACS analysis for surface expression of a pilin antigen that has homology to SpyM3_0104 identified in a different GAS serotype, M12. FACS analysis conducted with the SpyM3_0104 antisera was able to detect surface expression of the homologous SpyM3_0104 antigen on GAS serotype M12 2728. Table 30, below, quantitatively summarizes the FACS fluorescence values obtained for this GAS serotype in the presence of pre-immune antiserum, anti-SpyM3_0104 antiserum, and the difference in fluorescence value between the pre-immune and anti-SpyM3_0104 antiserum.

Table 30: Summary of FACS values for surface expression of SpyM3_0104 in an M12 serotype

2728		
Pre-immune	Anti-spyM3_0104	Change
198.57	288.75	90

Figure 84 provides the results of FACS analysis for surface expression of SPs_0106 on each of GAS serotypes M3 2721 and M3 3135. A shift in fluorescence is observed for each GAS serotype when anti-SPs_0106 antiserum is present, demonstrating cell surface expression. Table 31, below, quantitatively summarizes the FACS fluorescence values obtained for each GAS serotype in the presence of pre-immune antiserum, anti-SPs_0106 antiserum, and the difference in fluorescence value between the pre-immune and anti-SPs_0106 antiserum.

Table 31: Summary of FACS values for surface expression of SPs_0106 in M3 serotypes

2721			3135		
Pre-immune	Anti-SPs_0106	Change	Pre-immune	Anti-SPs_0106	Change
116	463.28	347	103.02	494.27	391

Figure 84 also provides the results of FACS analysis for surface expression of a pilin antigen that has homology to SPs_0106 identified in a different GAS serotype, M12. FACS analysis conducted with the SPs_0106 antisera was able to detect surface expression of the homologous SPs_0106 antigen on GAS serotype M12 2728. Table 32, below, quantitatively summarizes the FACS fluorescence values obtained for each GAS serotype in the presence of pre-immune antiserum, anti-SPs_0106 antiserum, and the difference in fluorescence value between the pre-immune and anti-SPs_0106 antiserum.

Table 32: Summary of FACS values for surface expression of SPs_0106 in an M12 serotype

2728		
Pre-immune	Anti-SPs_0106	Change
304.01	254.64	-49

(4) <u>Adhesin Island sequence within M12: GAS Adhesin Island 4 ("GAS AI-4")</u>

GAS Adhesin Island sequences within M12 serotype are outlined in Table 11 below. This GAS adhesin island 4 ("GAS AI-4") comprises surface proteins, a SrtC2 sortase, and a RofA regulatory protein.

GAS AI-4 surface proteins within may include a fimbrial protein, an F or F2 like fibronectin-binding protein, and a capsular polysaccharide adhesion protein (Cpa). GAS AI-4 surface proteins may also include a hypothetical surface protein in an open reading frame (orf). Preferably, each of these GAS AI-4 surface proteins include an LPXTG sortase substrate motif, such as LPXTG (SEQ ID NO: 122), VPXTG (SEQ ID NO: 137), QVXTG (SEQ ID NO: 138) or LPXAG (SEQ ID NO: 139).

GAS AI-4 includes a SrtC2 type sortase. GAS SrtC2 type sortases may preferably anchor surface proteins with a QVPTG (SEQ ID NO: 140) motif, particularly when the motif is followed by a hydrophobic region and a charged C terminus tail.

GAS AI-4 may also include a LepA putative signal peptidase I protein and a MsmRL protein.

Table 11: GAS AI-4 sequences from M12 isolate (A735)

AI-4 sequence identifier	Sortase substrate sequence or sortase type	Functional description
19224133		RofA regulatory protein
19224134	LPXTG	protein F
	SrtB	SrtB **(stop codon*)**
19224135	VPXTG	Cpa
19224136		LepA
19224137	QVXTG	EftLSL.A (fimbrial)
19224138	SrtC2	EftLSL.B
19224139	LPXAG	Orf2
19224140		MsmRL
19224141	LPXTG	protein F2

A schematic of AI-4 serotype M12 is shown in Figure 51A.

One of the open reading frames encodes a SrtC2-type sortase having an amino acid sequence nearly identical to the amino acid sequence of the SrtC2-type sortase of the AI-3 serotypes described above. See Figure 52B for an amino acid sequence alignment for each of the SrtC2 amino acid sequences.

Other proteins encoded by the open reading frames of the AI-4 serotype M12 are homologous to proteins encoded by other known adhesin islands in *S. pyogenes*, as well as the GAS AI-3 serotype M5 (Manfredo). Figure 52 is an amino acid alignment of the capsular polysaccharide adhesion protein (cpa) of AI-4 serotype M12 (19224135), GAS AI-3 serotype M5 (ORF78), *S. pyogenes* strain MGAS315 serotype M3 (21909634), *S. pyogenes* SSI-1 serotype M3 (28810257), *S. pyogenes* MGAS8232 serotype M3 (19745301), and GAS AI-2 serotype M1 (GAS15). The amino acid sequence of the AI-4 serotype M12 cpa shares a high degree of homology with other cpa proteins.

Figure 53 shows that the F-like fibronectin-binding protein encoded by the AI-4 serotype M12 open reading frame (19224134) shares homology with a F-like fibronectin-binding protein found in *S. pyogenes* strain MGAS10394 serotype M6 (50913503).

Figure 54 is an amino acid sequence alignment that illustrates that the F2-like fibronectin-binding protein of AI-4 serotype M12 (19224141) shares homology with the F2-like fibronectin-binding protein of *S. pyogenes* strain MGAS8232 serotype M3 (19745307), GAS AI-3 serotype M5 (ORF84), *S. pyogenes* strain SSI serotype M3 (28810263), and *S. pyogenes* strain MGAS315 serotype M3 (21909640).

Figure 55 is an amino acid sequence alignment that illustrates that the fimbrial protein of AI-4 serotype M12 (19224137) shares homology with the fimbrial protein of GAS AI-3 serotype M5 (ORF80), and the hypothetical protein of *S. pyogenes* strain MGAS315 serotype M3 (21909636), *S. pyogenes* strain SSI serotype M3 (28810259), *S. pyogenes* strain MGAS8732 serotype M3 (19745303), and *S. pyogenes* strain M1 GAS serotype M1 (13621428).

Figure 56 is an amino acid sequence alignment that illustrates that the hypothetical protein of GAS AI-4 serotype M12 (19224139) shares homology with the hypothetical protein of *S. pyogenes* strain MGAS315 serotype M3 (21909638), *S. pyogenes* strain SSI-1 serotype M3 (28810261), GAS AI-3 serotype M5 (ORF82), and *S. pyogenes* strain MGAS8232 serotype M3 (19745305).

The protein F2-like fibronectin-binding protein of the type 4 adhesin island also contains a highly conserved pilin motif and an E-box. Figure 60 indicates the amino acid sequence of the pilin motif and E-box in AI-4 serotype M12.

FACS analysis has confirmed that the GAS AI-4 surface proteins 19224134, 19224135, 19224137, and 19224141 are expressed on the surface of GAS. Figure 85 provides the results of FACS analysis for surface expression of 19224134 on GAS serotype M12 2728. A shift in fluorescence is observed when anti-19224134 antiserum is present, demonstrating cell surface expression. Table 33, below, quantitatively summarizes the FACS fluorescence values obtained for GAS serotype M12 2728 in the presence of pre-immune antiserum, anti-19224134 antiserum, and the difference in fluorescence value between the pre-immune and anti-19224134 antiserum.

Table 33: Summary of FACS values for surface expression of 19224134 in an M12 serotype

2728

Pre-immune	Anti-19224134	Change
137.8	485.32	348

Figure 85 also provides the results of FACS analysis for surface expression of a pilin antigen that has homology to 19224134 identified in a different GAS serotype, M6. FACS analysis conducted with the 19224134 antisera was able to detect surface expression of the homologous 19224134 antigen on each of GAS serotypes M6 2724, M6 3650, and M6 2894. Table 34, below, quantitatively summarizes the FACS fluorescence values obtained for each GAS serotype in the presence of pre-immune antiserum, anti-19224134 antiserum, and the difference in fluorescence value between the pre-immune and anti-19224134 antiserum.

Table 34: Summary of FACS values for surface expression of 19224134 in M6 serotypes

2724			3650			2894		
Pre-immune	Anti-19224134	Change	Pre-immune	Anti-19224134	Change	Pre-immune	Anti-19224134	Change
123.58	264.59	141	140.82	262.64	122	135.4	307.25	172

Figure 86 provides the results of FACS analysis for surface expression of 19224135 on GAS serotype M12 2728. A shift in fluorescence is observed when anti-19224135 antiserum is present, demonstrating cell surface expression. Table 35, below, quantitatively summarizes the FACS fluorescence values obtained for GAS serotype M12 2728 in the presence of pre-immune antiserum, anti-19224135 antiserum, and the difference in fluorescence value between the pre-immune and anti-19224135 antiserum.

Table 35: Summary of FACS values for surface expression of 19224135 in an M12 serotype

2728		
Pre-immune	Anti-19224135	Change
151.38	471.95	321

Figure 87 provides the results of FACS analysis for surface expression of 19224137 on GAS serotype M12 2728. A shift in fluorescence is observed when anti-19224137 antiserum is present, demonstrating cell surface expression. Table 36, below, quantitatively summarizes the FACS fluorescence values obtained for GAS serotype M12 2728 in the presence of pre-immune antiserum, anti-19224137 antiserum, and the difference in fluorescence value between the pre-immune and anti-19224137 antiserum.

Table 36: Summary of FACS values for surface expression of 19224137 in an M12 serotype

2728		
Pre-immune	Anti-19224137	Change
140.44	433.25	293

Figure 88 provides the results of FACS analysis for surface expression of 19224141 on GAS serotype M12 2728. A shift in fluorescence is observed when anti-19224141 antiserum is present, demonstrating cell surface expression. Table 37, below, quantitatively summarizes the FACS fluorescence values obtained for GAS serotype M12 2728 in the presence of pre-immune antiserum, anti-19224141 antiserum, and the difference in fluorescence value between the pre-immune and anti-19224141 antiserum.

Table 37: Summary of FACS values for surface expression of 19224141 in an M12 serotype

2728		
Pre-immune	Anti-19224141	Change
147.02	498	351

19224139 (designated as orf2) may also be expressed on the surface of GAS serotype M12 bacteria. Figure 175 shows the results of FACS analysis for surface expression of 19224139 on M12 strain 2728. A slight shift in fluorescence is observed, which demonstrates that some 19224139 may be expressed on the GAS cell surface.

Surface expression of 19224135 on M12 serotype GAS has also been confirmed by Western blot analysis. Figure 99 shows that while pre-immune sera (P α-4135) does not detect GAS M12 expression of 19224135, anti-19224135 immune sera (I α-4135) is able to detect 19224135 protein in both total GAS M12 extracts (M12 tot) and GAS M12 fractions enriched for cell surface proteins (M12 surf prot). The 19224135 proteins detected in the total GAS M12 extracts or the GAS M12 extracts enriched for surface proteins are also present as high molecular weight structures, indicating that 19224135 may be in an oligomeric (pilus) form. See also Figure 108, which provides a further Western blot showing that anti-19224135 antiserum (Anti-19224135) immunoreacts with high molecular weight structures in GAS M12 strain 2728 protein extracts enriched for surface proteins.

Surface expression of 19224137 on M12 serotype GAS has also been confirmed by Western blot analysis. Figure 100 shows that while pre-immune sera (P α-4137) does not detect GAS M12 expression of 19224137, anti-19224137 immune sera (I α-4137) is able to detect 19224137 protein in both total GAS M12 extracts (M12 tot) and GAS M12 fractions enriched for cell surface proteins (M12 surf prot). The 19224137 proteins detected in the total GAS M12 extracts or the GAS M12 extracts enriched for surface proteins are also present as high molecular weight structures, indicating that 19224137 may be in an oligomeric (pilus) form. See also Figure 108, which provides a further Western blot showing that anti-19224137 antiserum (Anti-19224137) immunoreacts with high molecular weight structures in GAS M12 strain 2728 protein extracts enriched for surface proteins.

Streptococcus pneumoniae

Adhesin island sequences can be identified in *Streptococcus pneumoniae* genomes. Several of these genomes include the publicly available *Streptococcus pneumoniae* TIGR4 genome or *Streptococcus pneumoniae* strain 670 genome. Examples of these *S. pneumoniae* AI sequence are discussed below.

S. pneumoniae Adhesin Islands generally include a series of open reading frames within a *S. pneumoniae* genome that encode for a collection of surface proteins and sortases. A *S. pneumoniae* Adhesin Island may encode for amino acid sequences comprising at least one surface protein. Alternatively, an *S. pneumoniae* Adhesin Island may encode for at least two surface proteins and at least one sortase. Preferably, a *S. pneumoniae* Adhesin Island encodes for at least three surface proteins and at least two sortases. One or more of the surface proteins may include an LPXTG motif (such as LPXTG (SEQ ID NO: 122)) or other sortase substrate motif. One or more *S. pneumoniae* AI surface proteins may participate in the formation of a pilus structure on the surface of the *S. pneumoniae* bacteria.

S. pneumoniae Adhesin Islands of the invention preferably include a divergently transcribed transcriptional regulator. The transcriptional regulator may regulate the expression of the *S. pneumoniae* AI operon.

The *S. pneumoniae* AI surface proteins may bind or otherwise adhere to fibrinogen, fibronectin, or collagen.

A schematic of the organization of a *S. pneumoniae* AI locus is provided in Figure 137. The locus comprises open reading frames encoding a transcriptional regulator (rlrA), cell wall surface proteins (rrgA, rrgB, rrgC), and sortases (srtB, srtC, srtD). Figure 137 also indicates the *S. pneumoniae* strain TIGR4 gene name corresponding to each of these open reading reading frames.

Tables 9 and 38 identify the genomic location of each of these open reading frames in *S. pneumoniae* strains TIGR4 and 670, respectively.

Table 9: S. pneumoniae AI sequences from TIGR4

Genomic Location	Strand	Length	PID	Synonym (AI Sequence Identifier)	Functional description
436302..437831	-	509	15900377	SP0461	transcriptional regulator
438326..441007	+	893	15900378	SP0462	cell wall surface anchor family protein
441231..443228	+	665	15900379	SP0463	cell wall surface anchor family protein
443275..444456	+	393	15900380	SP0464	cell wall surface anchor family protein
444675..444806	-	43	15900381	SP0465	hypothetical protein
444857..445696	+	279	15900382	SP0466	sortase
445791..446576	+	261	15900383	SP0467	sortase
446563..447414	+	283	15900384	SP0468	sortase

Table 38: S. pneumoniae strain 670 AI sequences

Genomic Location	Strand	AI Sequence Identifier\	Functional description
4383-5645	-	Orf1_670	IS1167, transposase
5910-7439	-	Orf2_670	transcriptional regulator, putative
7934-10606	+	Orf3_670	cell wall surface anchor family protein
10839-12773	+	Orf4_670	cell wall surface anchor family protein
12796-14001	+	Orf5_670	cell wall surface anchor family protein
14327-15241	+	Orf6_670	sortase, putative
15336-16121	+	Orf7_670	sortase, putative
16108-16959	+	Orf8_670	sortase, putative

The full-length nucleotide sequence of the *S. pneumoniae* strain 670 AI is also shown in Figure 101, as is its translated amino acid sequence.

At least eight other *S. pneumoniae* strains contain an adhesin island locus described by the locus depicted in Figure 137. These strains were identified by an amplification analysis. The genomes of different *S. pneumoniae* strains were amplified with eleven separate sets of primers. The sequence of each of these primers is provided below in Table 41.

Table 41: Sequences of primers used to amplify AI locus

Primer	Forward Primer Sequence	Reverse Primer Sequence

Pair		
1	ACTTTCTAATGAGTTGTTTAGGCG	AGCGACAAGCCACTGTATCATATT
2	CTGGTCGATAACTCCTTCAATCTT	GTACGACAAAAGTGTGGCTTGTT
3	GAATGCGATATTCAGGACCAACTA	ATCTCACTGAGTTAATCCGTTCAC
4	TGTATACAAGTGTGTCATTGCCAG	CATCTTCACCTGTTCTCACATTTT
5	GCGGTCTTTAGTCTTCAAAAACA	CAAGAGAAAAACACAGAGCCATAA
6	TTGCTTAAGTAAGAGAGAAAGGAGC	CAGGAGTATAGTGTCCGCTTTCTT
7	GGCAATGTTGACTTTATGAAGGTG	TATCAGCATCCCTTTATCTTCAAAC
8	TGAGATTTTCTCGTTTCTCTTAGC	AATAGACGATGGGTATTGATCATGT
9	CCGACGAACTTTGATGATTTATTG	ACCAACAGACGATGACTGTTAATC
10	AATGACTTTGAGCCTGTCTTGAT	TTCTACAATTTCCTGGCCATTATC
11	GCCATTTGGATCAGCTAAAAGTT	TTTTTCAACCCACTACAGTTGACA

These primers hybridized along the entire length of the AI locus to generate amplification products representative of sequences throughout the locus. See Figure 138, which is a schematic of the location where each of these primers hybridizes to the *S. pneumoniae* AI locus. Figure 139A provides the set of amplicons obtained from amplification of the AI locus in *S. pneumoniae* strain TIGR4. Figure 139B provides the length, in base pairs, of each amplicon in *S. pneumoniae* strain TIGR4. Amplification of the genome of *S. pneumoniae* strains 19A Hungary 6, 6B Finland 12, 6B Spain 2, 9V Spain 3, 14 CSR 10, 19F Taiwan 14, 23F Taiwan 15, and 23F Poland 16 produced a set of eleven amplicons for the eleven primer pairs, indicating that each of these strains also contained the *S. pneumoniae* AI locus.

The *S. pneumoniae* strains were also identified as containing the AI locus by comparative genome hybridization (CGH) analysis. The genomes of sixteen *S. pneumoniae* strains were interrogated for the presence of the AI locus by comparison to unique open reading frames of strain TIGR4. The AI locus was detected by this method in strains 19A Hungary 6 (19AHUN), 6B Finland 12 (6BFIN12), 6B Spain 2 (6BSP2), 14CSR10 (14 CSR10), 9V Spain 3 (9VSP3), 19F Taiwan 14 (19FTW14), 23F Taiwan 15 (19FTW15), and 23F Poland 16 (23FP16). See Figure 140.

The AI locus has been sequenced for each of these strains and the nucleotide and encoded amino acid seqeunce for each orf has been determined. An alignment of the complete nucleotide sequence of the adhesin island present in each of the ten strains is provided in Figure 196. Aligning the amino acid sequences encoded by the orfs reveals conservation of many of the AI polypeptide amino acid sequences. For example, Table 39 provides a comparison of the percent identities of the polypeptides encoded within the *S. pneumoniae* strain 670 and TIGR4 adhesin islands.

Table 39: Pecent identity comparison of S. pneumoniae strains AI sequences

S. pneumoniae strain 670 polypeptide	*S. pneumoniae* from TIGR4 polypeptide	Shared identity of polypeptides
Orf1_670	SP0460	99.3% identity in 422 aa overlap
Orf2_670	SP0461	100.0% identity in 509 aa overlap
Orf3_670	SP0462	83.2% identity in 895 aa overlap
Orf4_670	SP0463	47.9% identity in 678 aa overlap
Orf5_670	SP0464	99.7% identity in 393 aa overlap
Orf6_670	SP0466	100.0% identity in 279 aa overlap
Orf7_670	SP0467	94.2% identity in 260 aa overlap
Orf8_670	SP0468	91.5% identity in 283 aa overlap

Figures 141-147 each provide a multiple sequence alignment for the polypeptides encoded by one of the open reading frames in all ten AI-positive *S. pneumoniae* strains. In each of the sequence alignments, light shading indicates an LPXTG motif and dark shading indicates the presence of an E-box motif with the conserved glutamic acid residue of the E-box motif in bold.

The sequence alignments also revealed that the polypeptides encoded by most of the open reading frames may be divided into two groups of homology, *S. pneumoniae* AI-a and AI-b. *S. pneumoniae* strains that comprise AI-a include 14 CSR 10, 19A Hungary 6, 23F Poland 15, 670, 6B Finland 12, and 6B Spain 2. *S. pneumoniae* strains that comprise AI-b include 19F Taiwan 14, 9V Spain 3, 23F Taiwan 15, and TIGR4. An immunogenic composition of the invention may comprise one or more polypeptides from within each of *S. pneumoniae* AI-a and AI-b. For example, polypeptide RrgB, encoded by open reading frame 4, may be divided within two such groups of homology. One group contains the RrgB sequences of six *S. pneumoniae* strains and a second group contains the RrgB sequences of four *S. pneumoniae* strains. While the amino acid sequence of the strains within each individual group is 99-100 percent identical, the amino acid sequence identity of the strains in the first relative to the second group is only 48%. Table 41 provides the identity comparisons of the amino acid sequences encoded by each open reading frame for the ten *S. pneumoniae* strains.

Table 42: Conservation of amino acid sequences encoded by the *S. pneumoniae* AI locus

Putative Role of Polypeptide	Encoded by Orf	Groups of Homology	% Identity in Group	% Identity Between Groups
RlrA, transcriptional regulator	2	1 group (10 strains)	100	-
RrgA, cell wall surface protein	3	2 groups (6 + 4)	98-100	83
RrgB, cell wall surface protein	4	2 groups (6 + 4)	99-100	48
RrgC, cell wall surface protein	5	2 groups (6 + 4)	99-100	97
SrtB, putative sortase	6	2 groups (7 + 3)	99-100	97
SrtC, putative sortase	7	2 groups (6 + 4)	95-100	93
SrtD, putative sortase	8	2 groups (6 + 4)	99-100	92

The division of homology between the RrgB polypeptide in the *S. pneumoniae* strains is due a lack of amino acid sequence identity in the central amino acid residues. Amino acid residues 1-30 and 617-665 are identical for each of the ten *S. pneumoniae* strains. However, amino acid residues 31-616 share between 42 and 100 percent identity between strains. See Figure 149. The shared N- and C-terminal regions of identity in the RrgB polypeptides may be preferred portions of the RrgB polypeptide for use in an immunogenic composition. Similarly, shared regions of identity in any of the polypeptides encoded by the *S. pneumoniae* AI locus may be preferable for use in immunogenic compositions. One of skill in the art, using the amino acid alignments provided in Figures 141-147, would readily be able to determine these regions of identity.

The *S. pneumoniae* comprising these AI loci do, in fact, express high molecular weight polymers on their surface, indicating the presence of pili. See Figure 182, which shows detection of high molecular weight structures expressed by *S. pneumoniae* strains that comprise the adhesin island locus depicted in Figure 137; these strains are indicated as rlrA+. Confirming these findings, electron microscopy and negative staining detects the presence of pili extending from the surface of *S. pneumoniae*. See Figure 185. To demonstrate that the adhesin island locus was responsible for the pili, the rrgA-srtD region of TIGR 4 were deleted. Deletion of this region of the adhesin island resulted in a loss of pili expression. See Figure 186. See also Figure 235, which provides an electron micrograph of *S. pneumoniae* lacking the rrgA-srtD region immunogold stained using anti-RrgB and anti-RrgC antibodies. No pili can be seen. Similarly to that described above, a *S. pneumoniae* bacteria that lacks a transcriptional repressor, *mgrA*, of genes in the adhesin island expresses pili. See Figure 187. However, and as expected, a *S. pneumoniae* bacteria that lacks both the mgrA and adhesin island genes in the rrgA-srtD region does not express pili. See Figure 188.

These high molecular weight pili structures appear to play a role in adherence of *S. pneumoniae* to cells. *S. pneumoniae* TIGR4 that lack the pilus operon have significantly diminished ability to adhere to A549 alveolar cells in vitro. See Figure 184.

The Sp0463 (*S. pneumoniae* TIGR4 rrgB) adhesion island polypeptide is expressed in oligomeric form. Whole cell extracts were analyzed by Western blot using a Sp0463 antiserum. The antiserum cross-hybridized with high molecular weight Sp0463 polymers. See Figure 156. The antiserum did not cross-hybridize with polypeptides from D39 or R6 strains of *S. pneumoniae*, which do not contain the AI locus depicted in Figure 137. Immunogold labelling of *S. pneumoniae* TIGR 4 using RrgB antiserum confirms the presence of RrgB in pili. Figure 189 shows double-labeling of S. pneumoniae TIGR 4 bacteria with immunolabeling for RrgB (5 nm gold particles) and RrgC (10 nm gold particles) protein. The RrgB protein is detected as present at intervals along the pilus structure. The RrgC protein is detected at the tips of the pili. See Figure 234 at arrows; Figure 234 is a close up of a pilus in Figure 189 at the location indicated by *.

The RrgA protein appears to be present in and necessary for formation of high molecular weight structures on the surface of *S. pneumoniae* TIGR4. See Figure 181 which provides the results of Western blot analysis of TIGR4 *S. pneumoniae* lacking the gene encoding RrgA. No high molecular weight structures are detected in *S. pneumoniae* that do not express RrgA using antiserum raised against RrgB. See also Figure 183.

A detailed diagram of the amino acid sequence comparions of the RrgA protein in the ten *S.*. *pneumoniae* strains is shown in Figure 148. The diagram reveals the division of the individual *S. pneumoniae* strains into the two different homology groups.

The cell surface polypeptides encoded by the *S. pneumoniae* TIGR4 AI, Sp0462 (rrgA), Sp0463 (rrgB), and Sp0464 (rrgC), have been cloned and expressed. See examples 15-17. A polyacrylamide gel showing successful recombinant expression of RrgA is provided in Figure 190A. Detection of the RrgA protein, which is expressed in pET21b with a histidine tag, is also shown by Western blot analysis in Figure 190B, using an anti-histidine tag antibody.

Antibodies that detect RrgB and RrgC antibodies have been produced in mice. See Figures 191 and 192, which show detection of RrgB and RrgC, respectively, using the raised antibodies.

In addition to the identification of these *S. pneumoniae* adhesion islands, coding sequences for SrtB type sortases have been identified in several *S. pneumoniae* clinical isolates, demonstrating conservation of a SrtB type sortase across these isolates.

Recombinantly Produced AI polypeptides

It is also an aspect of the invention to alter a non-AI polypeptide to be expressed as an AI polypeptide. The non-AI polypeptide may be genetically manipulated to additionally contain AI polypeptide sequences, *e.g.*, a sortase substrate, pilin, or E-box motif, which may cause expression of the non-AI polypeptide as an AI polypeptide. Alternatively the non-AI polypeptide may be genetically manipulated to replace an amino acid sequence within the non-AI polypeptide for AI polypeptide sequences, *e.g.*, a sortase substrate, pilin, or E-box motif, which may cause expression of the non-AI polypeptide as an AI polypeptide. Any number of amino acid residues may be added to the non-AI polypeptide or may be replaced within the non-AI polypeptide to cause its expression as an AI polypeptide. At least 5, 6, 7, 8, 9, 10, 15, 20, 25, 30, 35, 50, 75, 100, 150, 200, or 250 amino acid residues may be replaced or added to the non-AI polypeptide amino acid sequence. GBS 322 may be one such non-AI polypeptide that may be expressed as an AI polypeptide.

GBS Adhesin Island Sequences

The GBS AI polypeptides of the invention can, of course, be prepared by various means (*e.g.* recombinant expression, purification from GBS, chemical synthesis *etc.*) and in various forms (*e.g.* native, fusions, glycosylated, non-glycosylated *etc.*). They are preferably prepared in substantially pure form (*i.e.* substantially free from other streptococcal or host cell proteins) or substantially isolated form.

The GBS AI proteins of the invention may include polypeptide sequences having sequence identity to the identified GBS proteins. The degree of sequence identity may vary depending on the amino acid sequence (a) in question, but is preferably greater than 50% (*e.g.* 60%, 65%, 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 99.5% or more). Polypeptides having sequence identity include homologs, orthologs, allelic variants and functional mutants of the identified GBS proteins. Typically, 50% identity or more between two proteins is considered to be an indication of functional equivalence. Identity between proteins is preferably determined by the Smith-Waterman homology search algorithm as implemented in the MPSRCH program (Oxford Molecular), using an affinity gap search with parameters *gap open penalty=12* and *gap extension penalty=1*.

The GBS adhesin island polynucleotide sequences may include polynucleotide sequences having sequence identity to the identified GBS adhesin island polynucleotide sequences. The degree of sequence identity may vary depending on the polynucleotide sequence in question, but is preferably greater than 50% (*e.g.* 60%, 65%, 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 99.5% or more).

The GBS adhesin island polynucleotide sequences of the invention may include polynucleotide fragments of the identified adhesin island sequences. The length of the fragment may vary depending on the polynucleotide sequence of the specific adhesin island sequence, but the fragment is preferably at least 10 consecutive polynucleotides, (*e.g.* at least 10, 12, 14, 16, 18, 20, 25, 30, 35, 40, 50, 60, 70, 80, 90, 100, 150, 200 or more).

The GBS adhesin island amino acid sequences of the invention may include polypeptide fragments of the identified GBS proteins. The length of the fragment may vary depending on the amino acid sequence of the specific GBS antigen, but the fragment is preferably at least 7 consecutive amino acids, (*e.g.* 8, 10, 12, 14, 16, 18, 20, 25, 30, 35, 40, 50, 60, 70, 80, 90, 100, 150, 200 or more). Preferably the fragment comprises one or more epitopes from the sequence. Other preferred fragments include (1) the N-terminal signal peptides of each identified GBS protein, (2) the identified GBS protein without their N-terminal signal peptides, and (3) each identified GBS protein wherein up to 10 amino acid residues (e.g. 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25 or more) are deleted from the N-terminus and/or the C-terminus *e.g.* the N-terminal amino acid residue may be deleted. Other fragments omit one or more domains of the protein (*e.g.* omission of a signal peptide, of a cytoplasmic domain, of a transmembrane domain, or of an extracellular domain).

GBS 80

Examples of preferred GBS 80 fragments are discussed below. Polynucleotide and polypeptide sequences of GBS 80 from a variety of GBS serotypes and strain isolates are set forth in Figures 18 and 22. The polynucleotide and polypeptide sequences for GBS 80 from GBS serotype V, strain isolate 2603 are also included below as SEQ ID NOS 1 and 2:

SEQ ID NO. 1

ATGAAATTATCGAAGAAGTTATTGTTTTCGGCTGCTGTTTTAACAATGGTGGCGGGGTCAACTGTTGAACCAGTA
GCTCAGTTTGCGACTGGAATGAGTATTGTAAGAGCTGCAGAAGTGTCACAAGAACGCCCAGCGAAAACAACAGTA
AATATCTATAAATTACAAGCTGATAGTTATAAATCGGAAATTACTTCTAATGGTGGTATCGAGAATAAAGACGGC
GAAGTAATATCTAACTATGCTAAACTTGGTGACAATGTAAAAGGTTTGCAAGGTGTACAGTTTAAACGTTATAAA
GTCAAGACGGATATTTCTGTTGATGAATTGAAAAAATTGACAACAGTTGAAGCAGCAGATGCAAAAGTTGGAACG
ATTCTTGAAGAAGGTGTCAGTCTACCTCAAAAAACTAATGCTCAAGGTTTGGTCGTCGATGCTCTGGATTCAAAA
AGTAATGTGAGATACTTGTATGTAGAAGATTTAAAGAATTCACCTTCAAACATTACCAAAGCTTATGCTGTACCG
TTTGTGTTGGAATTACCAGTTGCTAACTCTACAGGTACAGGTTTCCTTTCTGAAATTAATATTTACCCTAAAAAC
GTTGTAACTGATGAACCAAAAACAGATAAAGATGTTAAAAAATTAGGTCAGGACGATGCAGGTTATACGATTGGT
GAAGAATTCAAATGGTTCTTGAAATCTACAATCCCTGCCAATTTAGGTGACTATGAAAAATTTGAAATTACTGAT
AAATTTGCAGATGGCTTGACTTATAAATCTGTTGGAAAAATCAAGATTGGTTCGAAAACACTGAATAGAGATGAG
CACTACACTATTGATGAACCAACAGTTGATAACCAAAATACATTAAAAATTACGTTTAAACCAGAGAAATTTAAA
GAAATTGCTGAGCTACTTAAAGGAATGACCCTTGTTAAAAATCAAGATGCTCTTGATAAAGCTACTGCAAATACA
GATGATGCGGCATTTTTGGAAATTCCAGTTGCATCAACTATTAATGAAAAAGCAGTTTTAGGAAAAGCAATTGAA
AATACTTTTGAACTTCAATATGACCATACTCCTGATAAAGCTGACAATCCAAAACCATCTAATCCTCCAAGAAAA
CCAGAAGTTCATACTGGTGGGAAACGATTTGTAAAGAAAGACTCAACAGAAACACAAACACTAGGTGGTGCTGAG
TTTGATTTGTTGGCTTCTGATGGGACAGCAGTAAAATGGACAGATGCTCTTATTAAAGCGAATACTAATAAAAAC
TATATTGCTGGAGAAGCTGTTACTGGGCAACCAATCAAATTGAAATCACATACAGACGGTACGTTTGAGATTAAA
GGTTTGGCTTATGCAGTTGATGCGAATGCAGAGGGTACAGCAGTAACTTACAAATTAAAAGAAACAAAAGCACCA
GAAGGTTATGTAATCCCTGATAAAGAAATCGAGTTTACAGTATCACAAACATCTTATAATACAAAACCAACTGAC
ATCACGGTTGATAGTGCTGATGCAACACCTGATACAATTAAAAACAACAAACGTCCTTCAATCCCTAATACTGGT
GGTATTGGTACGGCTATCTTTGTCGCTATCGGTGCTGCGGTGATGGCTTTTGCTGTTAAGGGGATGAAGCGTCGT
ACAAAAGATAAC

SEQ ID NO: 2

MKLSKKLLFSAAVLTMVAGSTVEPVAQFATGMSIVRAAEVSQERPAKTTVNIYKLQADSYKSEITSNGGIENKDG
EVISNYAKLGDNVKGLQGVQFKRYKVKTDISVDELKKLTTVEAADAKVGTILEEGVSLPQKTNAQGLVVDALDSK
SNVRYLYVEDLKNSPSNITKAYAVPFVLELPVANSTGTGFLSEINIYPKNVVTDEPKTDKDVKKLGQDDAGYTIG
EEFKWFLKSTIPANLGDYEKFEITDKFADGLTYKSVGKIKIGSKTLNRDEHYTIDEPTVDNQNTLKITFKPEKFK

EIAELLKGMTLVKNQDALDKATANTDDAAFLEIPVASTINEKAVLGKAIENTFELQYDHTPDKADNPKPSNPPRK
PEVHTGGKRFVKKDSTETQTLGGAEFDLLASDGTAVKWTDALIKANTNKNYIAGEAVTGQPIKLKSHTDGTFEIK
GLAYAVDANAEGTAVTYKLKETKAPEGYVIPDKEIEFTVSQTSYNTKPTDITVDSADATPDTIKNNKRPS*IPNTG*
GIGTAIFVAIGAAVMAFAVKGMKRRTKDN

As described above, the compositions of the invention may include fragments of AI proteins. In some instances, removal of one or more domains, such as a leader or signal sequence region, a transmembrane region, a cytoplasmic region or a cell wall anchoring motif, may facilitate cloning of the gene encoding the protein and/or recombinant expression of the GBS AI protein. In addition, fragments comprising immunogenic epitopes of the cited GBS AI proteins may be used in the compositions of the invention.

For example, GBS 80 contains an N-terminal leader or signal sequence region which is indicated by the underlined sequence at the beginning of SEQ ID NO: 2 above. In one embodiment, one or more amino acids from the leader or signal sequence region of GBS 80 are removed. An example of such a GBS 80 fragment is set forth below as SEQ ID NO: 3:

SEQ ID NO: 3
AEVSQERPAKTTVNIYKLQADSYKSEITSNGGIENKDGEVISNYAKLGDNVKGLQGVQFKRYKVKTDISVDELKK
LTTVEAADAKVGTILEEGVSLPQKTNAQGLVVDALDSKSNVRYLYVEDLKNSPSNITKAYAVPFVLELPVANSTG
TGFLSEINIYPKNVVTDEPKTDKDVKKLGQDDAGYTIGEEFKWFLKSTIPANLGDYEKFEITDKFADGLTYKSVG
KIKIGSKTLNRDEHYTIDEPTVDNQNTLKITFKPEKFKEIAELLKGMTLVKNQDALDKATANTDDAAFLEIPVAS
TINEKAVLGKAIENTFELQYDHTPDKADNPKPSNPPRKPEVHTGGKRFVKKDSTETQTLGGAEFDLLASDGTAVK
WTDALIKANTNKNYIAGEAVTGQPIKLKSHTDGTFEIKGLAYAVDANAEGTAVTYKLKETKAPEGYVIPDKEIEF
TVSQTSYNTKPTDITVDSADATPDTIKNNKRPSIPNTGGIGTAIFVAIGAAVMAFAVKGMKRRTKDN

GBS 80 contains a C-terminal transmembrane region which is indicated by the underlined sequence near the end of SEQ ID NO: 2 above. In one embodiment, one or more amino acids from the transmembrane region and/or a cytoplasmic region are removed. An example of such a GBS 80 fragment is set forth below as SEQ ID NO: 4:

SEQ ID NO: 4
MKLSKKLLFSAAVLTMVAGSTVEPVAQFATGMSIVRAAEVSQERPAKTTVNIYKLQADSYKSEITSNGGIENKDG
EVISNYAKLGDNVKGLQGVQFKRYKVKTDISVDELKKLTTVEAADAKVGTILEEGVSLPQKTNAQGLVVDALDSK
SNVRYLYVEDLKNSPSNITKAYAVPFVLELPVANSTGTGFLSEINIYPKNVVTDEPKTDKDVKKLGQDDAGYTIG
EEFKWFLKSTIPANLGDYEKFEITDKFADGLTYKSVGKIKIGSKTLNRDEHYTIDEPTVDNQNTLKITFKPEKFK
EIAELLKGMTLVKNQDALDKATANTDDAAFLEIPVASTINEKAVLGKAIENTFELQYDHTPDKADNPKPSNPPRK
PEVHTGGKRFVKKDSTETQTLGGAEFDLLASDGTAVKWTDALIKANTNKNYIAGEAVTGQPIKLKSHTDGTFEIK
GLAYAVDANAEGTAVTYKLKETKAPEGYVIPDKEIEFTVSQTSYNTKPTDITVDSADATPDTIKNNKRPS*IPNTG*

GBS 80 contains an amino acid motif indicative of a cell wall anchor: **SEQ ID NO: 5** IPNTG (shown in italics in SEQ ID NO: 2 above). In some recombinant host cell systems, it may be preferable to remove this motif to facilitate secretion of a recombinant GBS 80 protein from the host cell. Accordingly, in one preferred fragment of GBS 80 for use in the invention, the transmembrane and/or cytoplasmic regions and the cell wall anchor motif are removed from GBS 80. An example of such a GBS 80 fragment is set forth below as SEQ ID NO: 6.

SEQ ID NO: 6
MKLSKKLLFSAAVLTMVAGSTVEPVAQFATGMSIVRAAEVSQERPAKTTVNIYKLQADSYKSEITSNGGIENKDG
EVISNYAKLGDNVKGLQGVQFKRYKVKTDISVDELKKLTTVEAADAKVGTILEEGVSLPQKTNAQGLVVDALDSK
SNVRYLYVEDLKNSPSNITKAYAVPFVLELPVANSTGTGFLSEINIYPKNVVTDEPKTDKDVKKLGQDDAGYTIG
EEFKWFLKSTIPANLGDYEKFEITDKFADGLTYKSVGKIKIGSKTLNRDEHYTIDEPTVDNQNTLKITFKPEKFK

EIAELLKGMTLVKNQDALDKATANTDDAAFLEIPVASTINEKAVLGKAIENTFELQYDHTPDKADNPKPSNPPRK
PEVHTGGKRFVKKDSTETQTLGGAEFDLLASDGTAVKWTDALIKANTNKNYIAGEAVTGQPIKLKSHTDGTFEIK
GLAYAVDANAEGTAVTYKLKETKAPEGYVIPDKEIEFTVSQTSYNTKPTDITVDSADATPDTIKNNKRPS

Alternatively, in some recombinant host cell systems, it may be preferable to use the cell wall anchor motif to anchor the recombinantly expressed protein to the cell wall. The extracellular domain of the expressed protein may be cleaved during purification or the recombinant protein may be left attached to either inactivated host cells or cell membranes in the final composition.

In one embodiment, the leader or signal sequence region, the transmembrane and cytoplasmic regions and the cell wall anchor motif are removed from the GBS 80 sequence. An example of such a GBS 80 fragment is set forth below as SEQ ID NO: 7.

SEQ ID NO: 7
AEVSQERPAKTTVNIYKLQADSYKSEITSNGGIENKDGEVISNYAKLGDNVKGLQGVQFKRYKVKTDISVDELKK
LTTVEAADAKVGTILEEGVSLPQKTNAQGLVVDALDSKSNVRYLYVEDLKNSPSNITKAYAVPFVLELPVANSTG
TGFLSEINIYPKNVVTDEPKTDKDVKKLGQDDAGYTIGEEFKWFLKSTIPANLGDYEKFEITDKFADGLTYKSVG
KIKIGSKTLNRDEHYTIDEPTVDNQNTLKITFKPEKFKEIAELLKGMTLVKNQDALDKATANTDDAAFLEIPVAS
TINEKAVLGKAIENTFELQYDHTPDKADNPKPSNPPRKPEVHTGGKRFVKKDSTETQTLGGAEFDLLASDGTAVK
WTDALIKANTNKNYIAGEAVTGQPIKLKSHTDGTFEIKGLAYAVDANAEGTAVTYKLKETKAPEGYVIPDKEIEF
TVSQTSYNTKPTDITVDSADATPDTIKNNKRPS

Applicants have identified a particularly immunogenic fragment of the GBS 80 protein. This immunogenic fragment is located towards the N-terminus of the protein and is underlined in the GBS 80 SEQ ID NO: 2 sequence below. The underlined fragment is set forth below as SEQ ID NO: 8.

SEQ ID NO: 2
MKLSKKLLFSAAVLTMVAGSTVEPVAQFATGMSIVRAAEVSQERPAKTTVNIYKLQADSYKSEITSNGGIENKDG
EVISNYAKLGDNVKGLQGVQFKRYKVKTDISVDELKKLTTVEAADAKVGTILEEGVSLPQKTNAQGLVVDALDSK
SNVRYLYVEDLKNSPSNITKAYAVPFVLELPVANSTGTGFLSEINIYPKNVVTDEPKTDKDVKKLGQDDAGYTIG
EEFKWFLKSTIPANLGDYEKFEITDKFADGLTYKSVGKIKIGSKTLNRDEHYTIDEPTVDNQNTLKITFKPEKFK
EIAELLKGMTLVKNQDALDKATANTDDAAFLEIPVASTINEKAVLGKAIENTFELQYDHTPDKADNPKPSNPPRK
PEVHTGGKRFVKKDSTETQTLGGAEFDLLASDGTAVKWTDALIKANTNKNYIAGEAVTGQPIKLKSHTDGTFEIK
GLAYAVDANAEGTAVTYKLKETKAPEGYVIPDKEIEFTVSQTSYNTKPTDITVDSADATPDTIKNNKRPS*IPNTG*
GIGTAIFVAIGAAVMAFAVKGMKRRTKDN

SEQ ID NO: 8
AEVSQERPAKTTVNIYKLQADSYKSEITSNGGIENKDGEVISNYAKLGDNVKGLQGVQFKRYKVKTDISVDELKK
LTTVEAADAKVGTILEEGVSLPQKTNAQGLVVDALDSKSNVRYLYVEDLKNSPSNITKAYAVPFVLELPVANSTG
TGFLSEINIYPKNVVTDEPKTDKDVKKLGQDDAGYTIGEEFKWFLKSTIPANLGDYEKFEITDKFADGLTYKSVG
KIKIGSKTLNRDEHYTIDEPTVDNQNTLKITFKPEKFKEIAELLKG

The immunogenicity of the protein encoded by SEQ ID NO: 7 was compared against PBS, GBS whole cell, GBS 80 (full length) and another fragment of GBS 80, located closer to the C-terminus of the peptide (SEQ ID NO: 9, below).

SEQ ID NO: 9
MTLVKNQDALDKATANTDDAAFLEIPVASTINEKAVLGKAIENTFELQYDHTPDKADNPKPSNPPRKPEVHTGGK
RFVKKDSTETQTLGGAEFDLLASDGTAVKWTDALIKANTNKNYIAGEAVTGQPIKLKSHTDGTFEIKGLAYAVDA
NAEGTAVTYKLKETKAPEGYVIPDKEIEFTVSQTSYNTKPTDITVDSADATPDTIKNNKRPS

Both an Active Maternal Immunization Assay and a Passive Maternal Immunization Assay were conducted on this collection of proteins.

As used herein, an Active Maternal Immunization assay refers to an *in vivo* protection assay where female mice are immunized with the test antigen composition. The female mice are then bred and their pups are challenged with a lethal dose of GBS. Serum titers of the female mice during the immunization schedule are measured as well as the survival time of the pups after challenge.

Specifically, the Active Maternal Immunization assays referred to herein used groups of four CD-1 female mice (Charles River Laboratories, Calco Italy). These mice were immunized intraperitoneally with the selected proteins in Freund's adjuvant at days 1, 21 and 35, prior to breeding. 6-8 weeks old mice received 20 μg protein/dose when immunized with a single antigen, 30-45 μg protein/dose (15 μg each antigen) when immunized with combination of antigens. The immune response of the dams was monitored by using serum samples taken on day 0 and 49. The female mice were bred 2-7 days after the last immunization (at approximately t= 36 – 37), and typically had a gestation period of 21 days. Within 48 hours of birth, the pups were challenged via I.P. with GBS in a dose approximately equal to a amount which would be sufficient to kill 70 – 90 % of unimmunized pups (as determined by empirical data gathered from PBS control groups). The GBS challenge dose is preferably administered in 50μl of THB medium. Preferably, the pup challenge takes place at 56 to 61 days after the first immunization. The challenge inocula were prepared starting from frozen cultures diluted to the appropriate concentration with THB prior to use. Survival of pups was monitored for 5 days after challenge.

As used herein, the Passive Maternal Immunization Assay refers to an *in vivo* protection assay where pregnant mice are passively immunized by injecting rabbit immune sera (or control sera) approximately 2 days before delivery. The pups are then challenged with a lethal dose of GBS.

Specifically, the Passive Maternal Immunization Assay referred to herein used groups of pregnant CD1 mice which were passively immunized by injecting 1 ml of rabbit immune sera or control sera via I.P., 2 days before delivery. Newborn mice (24-48 hrs after birth) are challenged via I.P. with a 70 - 90% lethal dose of GBS serotype III COH1. The challenge dose, obtained by diluting a frozen mid log phase culture, was administered in 50μl of THB medium.

For both assays, the number of pups surviving GBS infection was assessed every 12 hrs for 4 days. Statistical significance was estimated by Fisher's exact test.

The results of each assay for immunization with SEQ ID NO: 7, SEQ ID NO: 8, PBS and GBS whole cell are set forth in Tables 1 and 2 below.

TABLE 1: Immunization

Antigen	Alive/total	%Survival	Fisher's exact test
PBS (neg control)	13/80	16%	
GBS (whole cell)	54/65	83%	P<0.00000001
GBS80 (intact)	62/70	88%	P<0.00000001
GBS80 (fragment) SEQ ID 7	35/64	55%	P=0.0000013
GBS80 (fragment) SEQ ID 8	13/67	19%	P=0.66

Table 2: Passive Maternal Immunization

Antigen	Alive/total	%Survival	Fisher's exact test
PBS (neg control)	12/42	28%	
GBS (whole cell)	48/52	92%	P<0.00000001
GBS80 (intact)	48/55	87%	P<0.00000001
GBS80 (fragment) SEQ ID 7	45/57	79%	P=0.0000006
GBS80 (fragment) SEQ ID 8	13/54	24%	P=1

As shown in Tables 1 and 2, immunization with the SEQ ID NO: 7 GBS 80 fragment provided a substantially improved survival rate for the challenged pups than the comparison SEQ ID NO: 8 GBS 80 fragment. These results indicate that the SEQ ID NO: 7 GBS 80 fragment may comprise an important immunogenic epitope of GBS 80.

As discussed above, pilin motifs, containing conserved lysine (K) residues have been identified in GBS 80. The pilin motif sequences are underlined in SEQ ID NO: 2, below. Conserved lysine (K) residues are marked in bold, at amino acid residues 199 and 207 and at amino acid residues 368 and 375. The pilin sequences, in particular the conserved lysine residues, are thought to be important for the formation of oligomeric, pilus-like structures of GBS 80. Preferred fragments of GBS 80 include at least one conserved lysine residue. Preferably, fragments include at least one pilin sequence.

SEQ ID NO: 2

MKLSKKLLFSAAVLTMVAGSTVEPVAQFATGMSIVRAAEVSQERPAKTTVNIYKLQADSYKSEITSNGGIENKDG
EVISNYAKLGDNVKGLQGVQFKRYKVKTDISVDELKKLTTVEAADAKVGTILEEGVSLPQKTNAQGLVVDALDSK
SNVRYLYVEDLKNSPSNITKAYAVPFVLELPVANSTGTGFLSEINIYP**K**NVVTDEP**K**TDKDVKKLGQDDAGYTIG
EEFKWFLKSTIPANLGDYEKFEITDKFADGLTYKSVGKIKIGSKTLNRDEHYTIDEPTVDNQNTLKITFKPEKFK
EIAELLKGMTLVKNQDALDKATANTDDAAFLEIPVASTINEKAVLGKAIENTFELQYDHTPDKADNP**K**PSNPPR**K**
PEVHTGGKRFVKKDSTETQTLGGAEFDLLASDGTAVKWTDALIKANTNKNYIAGEAVTGQPIKLKSHTDGTFEIK
GLAYAVDANAEGTAVTYKLKETKAPEGYVIPDKEIEFTVSQTSYNTKPTDITVDSADATPDTIKNNKRPSIPNTG
GIGTAIFVAIGAAVMAFAVKGMKRRTKDN

E boxes containing conserved glutamic residues have also been identified in GBS 80. The E box motifs are underlined in SEQ ID NO: 2 below. The conserved glutamic acid (E) residues, at amino acid residues 392 and 471, are marked in bold. The E box motifs, in particular the conserved glutamic acid residues, are thought to be important for the formation of oligomeric pilus-like structures of GBS 80. Preferred fragments of GBS 80 include at least one conserved glutamic acid residue. Preferably, fragments include at least one E box motif.

SEQ ID NO: 2

MKLSKKLLFSAAVLTMVAGSTVEPVAQFATGMSIVRAAEVSQERPAKTTVNIYKLQADSYKSEITSNGGIENKDG
EVISNYAKLGDNVKGLQGVQFKRYKVKTDISVDELKKLTTVEAADAKVGTILEEGVSLPQKTNAQGLVVDALDSK
SNVRYLYVEDLKNSPSNITKAYAVPFVLELPVANSTGTGFLSEINIYPKNVVTDEPKTDKDVKKLGQDDAGYTIG
EEFKWFLKSTIPANLGDYEKFEITDKFADGLTYKSVGKIKIGSKTLNRDEHYTIDEPTVDNQNTLKITFKPEKFK
EIAELLKGMTLVKNQDALDKATANTDDAAFLEIPVASTINEKAVLGKAIENTFELQYDHTPDKADNPKPSNPPRK
PEVHTGGKRFVKKDST**E**TQTLGGAEFDLLASDGTAVKWTDALIKANTNKNYIAGEAVTGQPIKLKSHTDGTFEIK
GLAYAVDANAEGTAVTYKLK**E**TKAPEGYVIPDKEIEFTVSQTSYNTKPTDITVDSADATPDTIKNNKRPSIPNTG
GIGTAIFVAIGAAVMAFAVKGMKRRTKDN

GBS 104

Similarly, the following offers examples of preferred GBS 104 fragments. Nucleotide and amino acid sequences of GBS 104 sequenced from serotype V isolated strain 2603 are set forth below as SEQ ID NOS 10 and 11:

SEQ ID NO. 10

ATGAAAAAGAGACAAAAAATATGGAGAGGGTTATCAGTTACTTTACTAATCCTGTCCCAAATTCCATTTGGTATA
TTGGTACAAGGTGAAACCCAAGATACCAATCAAGCACTTGGAAAAGTAATTGTTAAAAAAACGGGAGACAATGCT
ACACCATTAGGCAAAGCGACTTTTGTGTTAAAAAATGACAATGATAAGTCAGAAACAAGTCACGAAACGGTAGAG
GGTTCTGGAGAAGCAACCTTTGAAAACATAAAACCTGGAGACTACACATTAAGAGAAGAAACAGCACCAATTGGT
TATAAAAAAACTGATAAAACCTGGAAAGTTAAAGTTGCAGATAACGGAGCAACAATAATCGAGGGTATGGATGCA
GATAAAGCAGAGAAACGAAAAGAAGTTTTGAATGCCCAATATCCAAAATCAGCTATTTATGAGGATACAAAAGAA
AATTACCCATTAGTTAATGTAGAGGGTTCCAAAGTTGGTGAACAATACAAAGCATTGAATCCAATAAATGGAAAA
GATGGTCGAAGAGAGATTGCTGAAGGTTGGTTATCAAAAAAAATTACAGGGGTCAATGATCTCGATAAGAATAAA
TATAAAATTGAATTAACTGTTGAGGGTAAAACCACTGTTGAAACGAAAGAACTTAATCAACCACTAGATGTCGTT
GTGCTATTAGATAATTCAAATAGTATGAATAATGAAAGAGCCAATAATTCTCAAAGAGCATTAAAAGCTGGGGAA
GCAGTTGAAAAGCTGATTGATAAAATTACATCAAATAAAGACAATAGAGTAGCTCTTGTGACATATGCCTCAACC
ATTTTTGATGGTACTGAAGCGACCGTATCAAAGGGAGTTGCCGATCAAAATGGTAAAGCGCTGAATGATAGTGTA
TCATGGGATTATCATAAAACTACTTTTACAGCAACTACACATAATTACAGTTATTTAAATTTAACAAATGATGCT
AACGAAGTTAATATTCTAAAGTCAAGAATTCCAAAGGAAGCGGAGCATATAAATGGGGATCGCACGCTCTATCAA
TTTGGTGCGACATTTACTCAAAAAGCTCTAATGAAAGCAAATGAAATTTTAGAGACACAAAGTTCTAATGCTAGA
AAAAAACTTATTTTTCACGTAACTGATGGTGTCCCTACGATGTCTTATGCCATAAATTTTAATCCTTATATATCA
ACATCTTACCAAAACCAGTTTAATTCTTTTTTAAATAAAATACCAGATAGAAGTGGTATTCTCCAAGAGGATTTT
ATAATCAATGGTGATGATTATCAAATAGTAAAAGGAGATGGAGAGAGTTTTAAACTGTTTTCGGATAGAAAAGTT
CCTGTTACTGGAGGAACGACACAAGCAGCTTATCGAGTACCGCAAAATCAACTCTCTGTAATGAGTAATGAGGGA
TATGCAATTAATAGTGGATATATTTATCTCTATTGGAGAGATTACAACTGGGTCTATCCATTTGATCCTAAGACA
AAGAAAGTTTCTGCAACGAAACAAATCAAAACTCATGGTGAGCCAACAACATTATACTTTAATGGAAATATAAGA
CCTAAAGGTTATGACATTTTTACTGTTGGGATTGGTGTAAACGGAGATCCTGGTGCAACTCCTCTTGAAGCTGAG
AAATTTATGCAATCAATATCAAGTAAAACAGAAAATTATACTAATGTTGATGATACAAATAAAATTTATGATGAG
CTAAATAAATACTTTAAAACAATTGTTGAGGAAAAACATTCTATTGTTGATGGAAATGTGACTGATCCTATGGGA
GAGATGATTGAATTCCAATTAAAAAATGGTCAAAGTTTTACACATGATGATTACGTTTTGGTTGGAAATGATGGC
AGTCAATTAAAAAATGGTGTGGCTCTTGGTGGACCAAACAGTGATGGGGGAATTTTAAAAGATGTTACAGTGACT
TATGATAAGACATCTCAAACCATCAAAATCAATCATTTGAACTTAGGAAGTGGACAAAAAGTAGTTCTTACCTAT
GATGTACGTTTAAAAGATAACTATATAAGTAACAAATTTTACAATACAAATAATCGTACAACGCTAAGTCCGAAG
AGTGAAAAAGAACCAAATACTATTCGTGATTTCCCAATTCCCAAAATTCGTGATGTTCGTGAGTTTCCGGTACTA
ACCATCAGTAATCAGAAGAAAATGGGTGAGGTTGAATTTATTAAAGTTAATAAAGACAAACATTCAGAATCGCTT
TTGGGAGCTAAGTTTCAACTTCAGATAGAAAAAGATTTTTCTGGGTATAAGCAATTTGTTCCAGAGGGAAGTGAT
GTTACAACAAAGAATGATGGTAAAATTTATTTTAAAGCACTTCAAGATGGTAACTATAAATTATATGAAATTTCA
AGTCCAGATGGCTATATAGAGGTTAAAACGAAACCTGTTGTGACATTTACAATTCAAAATGGAGAAGTTACGAAC
CTGAAAGCAGATCCAAATGCTAATAAAAATCAAATCGGGTATCTTGAAGGAAATGGTAAACATCTTATTACCAAC
ACTCCCAAACGCCCACCAGGTGTTTTTCCTAAAACAGGGGGAATTGGTACAATTGTCTATATATTAGTTGGTTCT
ACTTTTATGATACTTACCATTTGTTCTTTCCGTCGTAAACAATTG

SEQ ID NO. 11

MKKRQKIWRGLSVTLLILSQIPFGILVQGETQDTNQALGKVIVKKTGDNATPLGKATFVLKNDNDKSETSHETVE
GSGEATFENIKPGDYTLREETAPIGYKKTDKTWKVKVADNGATIIEGMDADKAEKRKEVLNAQYPKSAIYEDTKE
NYPLVNVEGSKVGEQYKALNPINGKDGRREIAEGWLSKKITGVNDLDKNKYKIELTVEGKTTVETKELNQPLDVV
VLLDNSNSMNNERANNSQRALKAGEAVEKLIDKITSNKDNRVALVTYASTIFDGTEATVSKGVADQNGKALNDSV
SWDYHKTTFTATTHNYSYLNLTNDANEVNILKSRIPKEAEHINGDRTLYQFGATFTQKALMKANEILETQSSNAR
KKLIFHVTDGVPTMSYAINFNPYISTSYQNQFNSFLNKIPDRSGILQEDFIINGDDYQIVKGDGESFKLFSDRKV
PVTGGTTQAAYRVPQNQLSVMSNEGYAINSGYIYLYWRDYNWVYPFDPKTKKVSATKQIKTHGEPTTLYFNGNIR
PKGYDIFTVGIGVNGDPGATPLEAEKFMQSISSKTENYTNVDDTNKIYDELNKYFKTIVEEKHSIVDGNVTDPMG
EMIEFQLKNGQSFTHDDYVLVGNDGSQLKNGVALGGPNSDGGILKDVTVTYDKTSQTIKINHLNLGSGQKVVLTY
DVRLKDNYISNKFYNTNNRTTLSPKSEKEPNTIRDFPIPKIRDVREFPVLTISNQKKMGEVEFIKVNKDKHSESL
LGAKFQLQIEKDFSGYKQFVPEGSDVTTKNDGKIYFKALQDGNYKLYEISSPDGYIEVKTKPVVTFTIQNGEVTN
LKADPNANKNQIGYLEGNGKHLITNTPKRPPGV*FPKTG*GIGTIVYILVGSTFMILTICSFRRKQL

GBS 104 contains an N-terminal leader or signal sequence region which is indicated by the underlined sequence at the beginning of SEQ ID NO 11 above. In one embodiment, one or more amino acid sequences from the leader or signal sequence region of GBS 104 are removed. An example of such a GBS 104 fragment is set forth below as SEQ ID NO 12.

SEQ ID NO 12
GETQDTNQALGKVIVKKTGDNATPLGKATFVLKNDNDKSETSHETVEGSGEATFENIKPGDYTLREETAPIGYKK
TDKTWKVKVADNGATIIEGMDADKAEKRKEVLNAQYPKSAIYEDTKENYPLVNVEGSKVGEQYKALNPINGKDGR
REIAEGWLSKKITGVNDLDKNKYKIELTVEGKTTVETKELNQPLDVVVLLDNSNSMNNERANNSQRALKAGEAVE
KLIDKITSNKDNRVALVTYASTIFDGTEATVSKGVADQNGKALNDSVSWDYHKTTFTATTHNYSYLNLTNDANEV
NILKSRIPKEAEHINGDRTLYQFGATFTQKALMKANEILETQSSNARKKLIFHVTDGVPTMSYAINFNPYISTSY
QNQFNSFLNKIPDRSGILQEDFIINGDDYQIVKGDGESFKLFSDRKVPVTGGTTQAAYRVPQNQLSVMSNEGYAI
NSGYIYLYWRDYNWVYPFDPKTKKVSATKQIKTHGEPTTLYFNGNIRPKGYDIFTVGIGVNGDPGATPLEAEKFM
QSISSKTENYTNVDDTNKIYDELNKYFKTIVEEKHSIVDGNVTDPMGEMIEFQLKNGQSFTHDDYVLVGNDGSQL
KNGVALGGPNSDGGILKDVTVTYDKTSQTIKINHLNLGSGQKVVLTYDVRLKDNYISNKFYNTNNRTTLSPKSEK
EPNTIRDFPIPKIRDVREFPVLTISNQKKMGEVEFIKVNKDKHSESLLGAKFQLQIEKDFSGYKQFVPEGSDVTT
KNDGKIYFKALQDGNYKLYEISSPDGYIEVKTKPVVTFTIQNGEVTNLKADPNANKNQIGYLEGNGKHLITNTPK
RPPGVFPKTGGIGTIVYILVGSTFMILTICSFRRKQL

GBS 104 contains a C-terminal transmembrane and/or cytoplasmic region which is indicated by the underlined region near the end of SEQ ID NO 11 above. In one embodiment, one or more amino acids from the transmembrane or cytomplasmic regions are removed. An example of such a GBS 104 fragment is set forth below as SEQ ID NO 13.

SEQ ID NO: 13
MKKRQKIWRGLSVTLLILSQIPFGILVQGETQDTNQALGKVIVKKTGDNATPLGKATFVLKNDNDKSETSHETVE
GSGEATFENIKPGDYTLREETAPIGYKKTDKTWKVKVADNGATIIEGMDADKAEKRKEVLNAQYPKSAIYEDTKE
NYPLVNVEGSKVGEQYKALNPINGKDGRREIAEGWLSKKITGVNDLDKNKYKIELTVEGKTTVETKELNQPLDVV
VLLDNSNSMNNERANNSQRALKAGEAVEKLIDKITSNKDNRVALVTYASTIFDGTEATVSKGVADQNGKALNDSV
SWDYHKTTFTATTHNYSYLNLTNDANEVNILKSRIPKEAEHINGDRTLYQFGATFTQKALMKANEILETQSSNAR
KKLIFHVTDGVPTMSYAINFNPYISTSYQNQFNSFLNKIPDRSGILQEDFIINGDDYQIVKGDGESFKLFSDRKV
PVTGGTTQAAYRVPQNQLSVMSNEGYAINSGYIYLYWRDYNWVYPFDPKTKKVSATKQIKTHGEPTTLYFNGNIR
PKGYDIFTVGIGVNGDPGATPLEAEKFMQSISSKTENYTNVDDTNKIYDELNKYFKTIVEEKHSIVDGNVTDPMG
EMIEFQLKNGQSFTHDDYVLVGNDGSQLKNGVALGGPNSDGGILKDVTVTYDKTSQTIKINHLNLGSGQKVVLTY
DVRLKDNYISNKFYNTNNRTTLSPKSEKEPNTIRDFPIPKIRDVREFPVLTISNQKKMGEVEFIKVNKDKHSESL
LGAKFQLQIEKDFSGYKQFVPEGSDVTTKNDGKIYFKALQDGNYKLYEISSPDGYIEVKTKPVVTFTIQNGEVTN
LKADPNANKNQIGYLEGNGKHLITNT

In one embodiment, one or more amino acids from the leader or signal sequence region and one or more amino acids from the transmembrane or cytoplasmic regions are removed. An example of such a GBS 104 fragment is set forth below as SEQ ID NO 14.

SEQ ID NO: 14
GETQDTNQALGKVIVKKTGDNATPLGKATFVLKNDNDKSETSHETVEGSGEATFENIKPGDYTLREETAPIGYKK
TDKTWKVKVADNGATIIEGMDADKAEKRKEVLNAQYPKSAIYEDTKENYPLVNVEGSKVGEQYKALNPINGKDGR
REIAEGWLSKKITGVNDLDKNKYKIELTVEGKTTVETKELNQPLDVVVLLDNSNSMNNERANNSQRALKAGEAVE
KLIDKITSNKDNRVALVTYASTIFDGTEATVSKGVADQNGKALNDSVSWDYHKTTFTATTHNYSYLNLTNDANEV
NILKSRIPKEAEHINGDRTLYQFGATFTQKALMKANEILETQSSNARKKLIFHVTDGVPTMSYAINFNPYISTSY
QNQFNSFLNKIPDRSGILQEDFIINGDDYQIVKGDGESFKLFSDRKVPVTGGTTQAAYRVPQNQLSVMSNEGYAI
NSGYIYLYWRDYNWVYPFDPKTKKVSATKQIKTHGEPTTLYFNGNIRPKGYDIFTVGIGVNGDPGATPLEAEKFM
QSISSKTENYTNVDDTNKIYDELNKYFKTIVEEKHSIVDGNVTDPMGEMIEFQLKNGQSFTHDDYVLVGNDGSQL
KNGVALGGPNSDGGILKDVTVTYDKTSQTIKINHLNLGSGQKVVLTYDVRLKDNYISNKFYNTNNRTTLSPKSEK
EPNTIRDFPIPKIRDVREFPVLTISNQKKMGEVEFIKVNKDKHSESLLGAKFQLQIEKDFSGYKQFVPEGSDVTT
KNDGKIYFKALQDGNYKLYEISSPDGYIEVKTKPVVTFTIQNGEVTNLKADPNANKNQIGYLEGNGKHLITNT

GBS 104, like GBS 80, contains an amino acid motif indicative of a cell wall anchor: **SEQ ID NO: 123** FPKTG (shown in italics in SEQ ID NO: 11 above). In some recombinant host cell systems, it may be preferable to remove this motif to facilitate secretion of a recombinant GBS 104 protein from the host cell. Accordingly, in one preferred fragment of GBS 104 for use in the invention, only the transmembrane and/or cytoplasmic regions and the cell wall anchor motif are removed from GBS 104. Alternatively, in some recombinant host cell systems, it may be preferable to use the cell wall anchor motif to anchor the recombinantly expressed protein to the cell wall. The extracellular domain of the expressed protein may be cleaved during purification or the recombinant protein may be left attached to either inactivated host cells or cell membranes in the final composition.

Two pilin motifs, containing conserved lysine (K) residues, have been identified in GBS 104. The pilin motif sequences are underlined in SEQ ID NO: 11, below. Conserved lysine (K) residues are marked in bold, at amino acid residues 141 and 149 and at amino acid residues 499 and 507. The pilin sequence, in particular the conserved lysine residues, are thought to be important for the formation of oligomeric, pilus-like structures of GBS 104. Preferred fragments of GBS 104 include at least one conserved lysine residue. Preferably, fragments include at least one pilin sequence.

SEQ ID NO. 11

MKKRQKIWRGLSVTLLILSQIPFGILVQGETQDTNQALGKVIVKKTGDNATPLGKATFVLKNDNDKSETSHETVE
GSGEATFENIKPGDYTLREETAPIGYKKTDKTWKVKVADNGATIIEGMDADKAEKRKEVLNAQYP**K**SAIYEDT**K**E
NYPLVNVEGSKVGEQYKALNPINGKDGRREIAEGWLSKKITGVNDLDKNKYKIELTVEGKTTVETKELNQPLDVV
VLLDNSNSMNNERANNSQRALKAGEAVEKLIDKITSNKDNRVALVTYASTIFDGTEATVSKGVADQNGKALNDSV
SWDYHKTTFTATTHNYSYLNLTNDANEVNILKSRIPKEAEHINGDRTLYQFGATFTQKALMKANEILETQSSNAR
KKLIFHVTDGVPTMSYAINFNPYISTSYQNQFNSFLNKIPDRSGILQEDFIINGDDYQIVKGDGESFKLFSDRKV
PVTGGTTQAAYRVPQNQLSVMSNEGYAINSGYIYLYWRDYNWVYPFDP**K**TKKVSAT**K**QIKTHGEPTTLYFNGNIR
PKGYDIFTVGIGVNGDPGATPLEAEKFMQSISSKTENYTNVDDTNKIYDELNKYFKTIVEEKHSIVDGNVTDPMG
EMIEFQLKNGQSFTHDDYVLVGNDGSQLKNGVALGGPNSDGGILKDVTVTYDKTSQTIKINHLNLGSGQKVVLTY
DVRLKDNYISNKFYNTNNRTTLSPKSEKEPNTIRDFPIPKIRDVREFPVLTISNQKKMGEVEFIKVNKDKHSESL
LGAKFQLQIEKDFSGYKQFVPEGSDVTTKNDGKIYFKALQDGNYKLYEISSPDGYIEVKTKPVVTFTIQNGEVTN
LKADPNANKNQIGYLEGNGKHLITNTPKRPPGVFPKTGGIGTIVYILVGSTFMILTICSFRRKQL

Two E boxes containing a conserved glutamic residues have also been identified in GBS 104. The E box motifs are underlined in SEQ ID NO: 11 below. The conserved glutamic acid (E) residues, at amino acid residues 94 and 798, are marked in bold. The E box motifs, in particular the conserved glutamic acid residues, are thought to be important for the formation of oligomeric pilus-like structures of GBS 104. Preferred fragments of GBS 104 include at least one conserved glutamic acid residue. Preferably, fragments include at least one E box motif.

SEQ ID NO. 11

MKKRQKIWRGLSVTLLILSQIPFGILVQGETQDTNQALGKVIVKKTGDNATPLGKATFVLKNDNDKSETSHETVE
GSGEATFENIKPGDYTLR**E**ETAPIGYKKTDKTWKVKVADNGATIIEGMDADKAEKRKEVLNAQYPKSAIYEDTKE
NYPLVNVEGSKVGEQYKALNPINGKDGRREIAEGWLSKKITGVNDLDKNKYKIELTVEGKTTVETKELNQPLDVV
VLLDNSNSMNNERANNSQRALKAGEAVEKLIDKITSNKDNRVALVTYASTIFDGTEATVSKGVADQNGKALNDSV
SWDYHKTTFTATTHNYSYLNLTNDANEVNILKSRIPKEAEHINGDRTLYQFGATFTQKALMKANEILETQSSNAR
KKLIFHVTDGVPTMSYAINFNPYISTSYQNQFNSFLNKIPDRSGILQEDFIINGDDYQIVKGDGESFKLFSDRKV
PVTGGTTQAAYRVPQNQLSVMSNEGYAINSGYIYLYWRDYNWVYPFDPKTKKVSATKQIKTHGEPTTLYFNGNIR
PKGYDIFTVGIGVNGDPGATPLEAEKFMQSISSKTENYTNVDDTNKIYDELNKYFKTIVEEKHSIVDGNVTDPMG
EMIEFQLKNGQSFTHDDYVLVGNDGSQLKNGVALGGPNSDGGILKDVTVTYDKTSQTIKINHLNLGSGQKVVLTY
DVRLKDNYISNKFYNTNNRTTLSPKSEKEPNTIRDFPIPKIRDVREFPVLTISNQKKMGEVEFIKVNKDKHSESL
LGAKFQLQIEKDFSGYKQFVPEGSDVTTKNDGKIYFKALQDGNYKLY**E**ISSPDGYIEVKTKPVVTFTIQNGEVTN
LKADPNANKNQIGYLEGNGKHLITNTPKRPPGVFPKTGGIGTIVYILVGSTFMILTICSFRRKQL

GBS 067

The following offers examples of preferred GBS 067 fragments. Nucleotide and amino acid sequence of GBS 067 sequences from serotype V isolated strain 2603 are set forth below as SEQ ID NOS: 15 and 16.

SEQ ID NO: 15

ATGAGAAAATACCAAAAATTTTCTAAAATATTGACGTTAAGTCTTTTTTGTTTGTCGCAAATACCGCTTAATACC
AATGTTTTAGGGGAAAGTACCGTACCGGAAAATGGTGCTAAAGGAAAGTTAGTTGTTAAAAAGACAGATGACCAG
AACAAACCACTTTCAAAAGCTACCTTTGTTTTAAAAACTACTGCTCATCCAGAAAGTAAAATAGAAAAAGTAACT
GCTGAGCTAACAGGTGAAGCTACTTTTGATAATCTCATACCTGGAGATTATACTTTATCAGAAGAAACAGCGCCC
GAAGGTTATAAAAAGACTAACCAGACTTGGCAAGTTAAGGTTGAGAGTAATGGAAAAACTACGATACAAAATAGT
GGTGATAAAAATTCCACAATTGGACAAAATCAGGAAGAACTAGATAAGCAGTATCCCCCCACAGGAATTTATGAA
GATACAAAGGAATCTTATAAACTTGAGCATGTTAAAGGTTCAGTTCCAAATGGAAAGTCAGAGGCAAAAGCAGTT
AACCCATATTCAAGTGAAGGTGAGCATATAAGAGAAATTCCAGAGGGAACATTATCTAAACGTATTTCAGAAGTA
GGTGATTTAGCTCATAATAAATATAAAATTGAGTTAACTGTCAGTGGAAAAACCATAGTAAAACCAGTGGACAAA
CAAAAGCCGTTAGATGTTGTCTTCGTACTCGATAATTCTAACTCAATGAATAACGATGGCCCAAATTTTCAAAGG
CATAATAAAGCCAAGAAAGCTGCCGAAGCTCTTGGGACCGCAGTAAAAGATATTTTAGGAGCAAACAGTGATAAT
AGGGTTGCATTAGTTACCTATGGTTCAGATATTTTTGATGGTAGGAGTGTAGATGTCGTAAAAGGATTTAAAGAA
GATGATAAATATTATGGCCTTCAAACTAAGTTCACAATTCAGACAGAGAATTATAGTCATAAACAATTAACAAAT
AATGCTGAAGAGATTATAAAAAGGATTCCGACAGAAGCTCCTAAAGCTAAGTGGGGATCTACTACCAATGGATTA
ACTCCAGAGCAACAAAAGGAGTACTATCTTAGTAAAGTAGGAGAAACATTTACTATGAAAGCCTTCATGGAGGCA
GATGATATTTTGAGTCAAGTAAATCGAAATAGTCAAAAAATTATTGTTCATGTAACTGATGGTGTTCCTACGAGA
TCATATGCTATTAATAATTTTAAACTGGGTGCATCATATGAAAGCCAATTTGAACAAATGAAAAAAAATGGATAT
CTAAATAAAAGTAATTTTCTACTTACTGATAAGCCCGAGGATATAAAAGGAAATGGGGAGAGTTACTTTTTGTTT
CCCTTAGATAGTTATCAAACACAGATAATCTCTGGAAACTTACAAAAACTTCATTATTTAGATTTAAATCTTAAT
TACCCTAAAGGTACAATTTATCGAAATGGACCAGTGAAAGAACATGGAACACCAACCAAACTTTATATAAATAGT
TTAAAACAGAAAAATTATGACATTTTTAATTTTGGTATCGATATATCTGGTTTTAGACAAGTTTATAATGAGGAG
TATAAGAAAAATCAAGATGGTACTTTTCAAAAATTGAAAGAGGAAGCTTTTAAACTTTCAGATGGAGAAATCACA
GAACTAATGAGGTCGTTCTCTTCCAAACCTGAGTACTACACCCCTATCGTAACTTCAGCCGATACATCTAACAAT
GAAATTTTATCTAAAATTCAGCAACAATTTGAAACGATTTTAACAAAAGAAAACTCAATTGTTAATGGAACTATC
GAAGATCCTATGGGTGATAAAATCAATTTACAGCTTGGTAATGGACAAACATTACAGCCAAGTGATTATACTTTA
CAGGGAAATGATGGAAGTGTAATGAAGGATGGTATTGCAACTGGTGGGCCTAATAATGATGGTGGAATACTTAAG
GGGGTTAAATTAGAATACATCGGAAATAAACTCTATGTTAGAGGTTTGAATTTAGGAGAAGGTCAAAAAGTAACA
CTCACATATGATGTGAAACTAGATGACAGTTTTATAAGTAACAAATTCTATGACACTAATGGTAGAACAACATTG
AATCCTAAGTCAGAGGATCCTAATACACTTAGAGATTTTCCAATCCCTAAAATTCGTGATGTGAGAGAATATCCT
ACAATAACGATTAAAAACGAGAAGAAGTTAGGTGAAATTGAATTTATAAAAGTTGATAAAGATAATAATAAGTTG
CTTCTCAAAGGAGCTACGTTTGAACTTCAAGAATTTAATGAAGATTATAAACTTTATTTACCAATAAAAAATAAT
AATTCAAAAGTAGTGACGGGAGAAAACGGCAAAATTTCTTACAAAGATTTGAAAGATGGCAAATATCAGTTAATA
GAAGCAGTTTCGCCGGAGGATTATCAAAAAATTACTAATAAACCAATTTTAACTTTTGAAGTGGTTAAAGGATCG
ATAAAAAATATAATAGCTGTTAATAAACAGATTTCTGAATATCATGAGGAAGGTGACAAGCATTTAATTACCAAC
ACGCATATTCCACCAAAAGGAATTATTCCTATGACAGGTGGGAAAGGAATTCTATCTTTCATTTTAATAGGTGGA
GCTATGATGTCTATTGCAGGTGGAATTTATATTTGGAAAAGGTATAAGAAATCTAGTGATATGTCCATCAAAAAA
GAT

SEQ ID NO: 16

MRKYQKFSKILTLSLFCLSQIPLNTNVLGESTVPENGAKGKLVVKKTDDQNKPLSKATFVLKTTAHPESKIEKVT
AELTGEATFDNLIPGDYTLSEETAPEGYKKTNQTWQVKVESNGKTTIQNSGDKNSTIGQNQEELDKQYPPTGIYE
DTKESYKLEHVKGSVPNGKSEAKAVNPYSSEGEHIREIPEGTLSKRISEVGDLAHNKYKIELTVSGKTIVKPVDK
QKPLDVVFVLDNSNSMNNDGPNFQRHNKAKKAAEALGTAVKDILGANSDNRVALVTYGSDIFDGRSVDVVKGFKE
DDKYYGLQTKFTIQTENYSHKQLTNNAEEIIKRIPTEAPKAKWGSTTNGLTPEQQKEYYLSKVGETFTMKAFMEA
DDILSQVNRNSQKIIVHVTDGVPTRSYAINNFKLGASYESQFEQMKKNGYLNKSNFLLTDKPEDIKGNGESYFLF
PLDSYQTQIISGNLQKLHYLDLNLNYPKGTIYRNGPVKEHGTPTKLYINSLKQKNYDIFNFGIDISGFRQVYNEE
YKKNQDGTFQKLKEEAFKLSDGEITELMRSFSSKPEYYTPIVTSADTSNNEILSKIQQQFETILTKENSIVNGTI
EDPMGDKINLQLGNGQTLQPSDYTLQGNDGSVMKDGIATGGPNNDGGILKGVKLEYIGNKLYVRGLNLGEGQKVT
LTYDVKLDDSFISNKFYDTNGRTTLNPKSEDPNTLRDFPIPKIRDVREYPTITIKNEKKLGEIEFIKVDKDNNKL
LLKGATFELQEFNEDYKLYLPIKNNNSKVVTGENGKISYKDLKDGKYQLIEAVSPEDYQKITNKPILTFEVVKGS
IKNIIAVNKQISEYHEEGDKHLITNTHIPPKGI*IPMTG*GKGILSFILIGGAMMSIAGGIYIWKRYKKSSDMSIKK
D

GBS 067 contains a C-terminus transmembrane region which is indicated by the underlined region closest to the C-terminus of SEQ ID NO: 16 above. In one embodiment, one or more amino acids from the transmembrane region is removed and or the amino acid is truncated before the transmembrane region. An example of such a GBS 067 fragment is set forth below as SEQ ID NO: 17.

SEQ ID NO: 17

MRKYQKFSKILTLSLFCLSQIPLNTNVLGESTVPENGAKGKLVVKKTDDQNKPLSKATFVLKTTAHPESKIEKVT
AELTGEATFDNLIPGDYTLSEETAPEGYKKTNQTWQVKVESNGKTTIQNSGDKNSTIGQNQEELDKQYPPTGIYE
DTKESYKLEHVKGSVPNGKSEAKAVNPYSSEGEHIREIPEGTLSKRISEVGDLAHNKYKIELTVSGKTIVKPVDK
QKPLDVVFVLDNSNSMNNDGPNFQRHNKAKKAAEALGTAVKDILGANSDNRVALVTYGSDIFDGRSVDVVKGFKE
DDKYYGLQTKFTIQTENYSHKQLTNNAEEIIKRIPTEAPKAKWGSTTNGLTPEQQKEYYLSKVGETFTMKAFMEA
DDILSQVNRNSQKIIVHVTDGVPTRSYAINNFKLGASYESQFEQMKKNGYLNKSNFLLTDKPEDIKGNGESYFLF
PLDSYQTQIISGNLQKLHYLDLNLNYPKGTIYRNGPVKEHGTPTKLYINSLKQKNYDIFNFGIDISGFRQVYNEE
YKKNQDGTFQKLKEEAFKLSDGEITELMRSFSSKPEYYTPIVTSADTSNNEILSKIQQQFETILTKENSIVNGTI
EDPMGDKINLQLGNGQTLQPSDYTLQGNDGSVMKDGIATGGPNNDGGILKGVKLEYIGNKLYVRGLNLGEGQKVT
LTYDVKLDDSFISNKFYDTNGRTTLNPKSEDPNTLRDFPIPKIRDVREYPTITIKNEKKLGEIEFIKVDKDNNKL
LLKGATFELQEFNEDYKLYLPIKNNNSKVVTGENGKISYKDLKDGKYQLIEAVSPEDYQKITNKPILTFEVVKGS
IKNIIAVNKQISEYHEEGDKHLITNTHIPPKGIIPMTGGKGILS

GBS 067 contains an amino acid motif indicative of a cell wall anchor (an LPXTG (SEQ ID NO: 122) motif): **SEQ ID NO: 18** IPMTG. (shown in italics in SEQ ID NO: 16 above). In some recombinant host cell systems, it may be preferable to remove this motif to facilitate secretion of a recombinant GBS 067 protein from the host cell. Accordingly, in one preferred fragment of GBS 067 for use in the invention, the transmembrane and the cell wall anchor motif are removed from GBS 67. An example of such a GBS 067 fragment is set forth below as SEQ ID NO: 19.

SEQ ID NO: 19

MRKYQKFSKILTLSLFCLSQIPLNTNVLGESTVPENGAKGKLVVKKTDDQNKPLSKATFVLKTTAHPESKIEKVT
AELTGEATFDNLIPGDYTLSEETAPEGYKKTNQTWQVKVESNGKTTIQNSGDKNSTIGQNQEELDKQYPPTGIYE
DTKESYKLEHVKGSVPNGKSEAKAVNPYSSEGEHIREIPEGTLSKRISEVGDLAHNKYKIELTVSGKTIVKPVDK
QKPLDVVFVLDNSNSMNNDGPNFQRHNKAKKAAEALGTAVKDILGANSDNRVALVTYGSDIFDGRSVDVVKGFKE
DDKYYGLQTKFTIQTENYSHKQLTNNAEEIIKRIPTEAPKAKWGSTTNGLTPEQQKEYYLSKVGETFTMKAFMEA
DDILSQVNRNSQKIIVHVTDGVPTRSYAINNFKLGASYESQFEQMKKNGYLNKSNFLLTDKPEDIKGNGESYFLF
PLDSYQTQIISGNLQKLHYLDLNLNYPKGTIYRNGPVKEHGTPTKLYINSLKQKNYDIFNFGIDISGFRQVYNEE
YKKNQDGTFQKLKEEAFKLSDGEITELMRSFSSKPEYYTPIVTSADTSNNEILSKIQQQFETILTKENSIVNGTI
EDPMGDKINLQLGNGQTLQPSDYTLQGNDGSVMKDGIATGGPNNDGGILKGVKLEYIGNKLYVRGLNLGEGQKVT
LTYDVKLDDSFISNKFYDTNGRTTLNPKSEDPNTLRDFPIPKIRDVREYPTITIKNEKKLGEIEFIKVDKDNNKL
LLKGATFELQEFNEDYKLYLPIKNNNSKVVTGENGKISYKDLKDGKYQLIEAVSPEDYQKITNKPILTFEVVKGS
IKNIIAVNKQISEYHEEGDKHLITNTHIPPKGI

Alternatively, in some recombinant host cell systems, it may be preferable to use the cell wall anchor motif to anchor the recombinantly expressed protein to the cell wall. The extracellular domain of the expressed protein may be cleaved during purification or the recombinant protein may be left attached to either inactivated host cells or cell membranes in the final composition.

Three pilin motifs, containing conserved lysine (K) residues have been identified in GBS 67. The pilin motif sequences are underlined in SEQ ID NO: 16, below. Conserved lysine (K) residues are marked in bold, at amino acid residues 478 and 488, at amino acid residues 340 and 342, and at amino acid residues 703 and 717. The pilin sequences, in particular the conserved lysine residues, are thought to be important for the formation of oligomeric, pilus-like structures of GBS 67. Preferred fragments of GBS 67 include at least one conserved lysine residue. Preferably, fragments include at least one pilin sequence.

SEQ ID NO: 16

MRKYQKFSKILTLSLFCLSQIPLNTNVLGESTVPENGAKGKLVVKKTDDQNKPLSKATFVLKTTAHPESKIEKVT
AELTGEATFDNLIPGDYTLSEETAPEGYKKTNQTWQVKVESNGKTTIQNSGDKNSTIGQNQEELDKQYPPTGIYE
DTKESYKLEHVKGSVPNGKSEAKAVNPYSSEGEHIREIPEGTLSKRISEVGDLAHNKYKIELTVSGKTIVKPVDK
QKPLDVVFVLDNSNSMNNDGPNFQRHNKAKKAAEALGTAVKDILGANSDNRVALVTYGSDIFDGRSVDVVKGFKE
DDKYYGLQTKFTIQTENYSHKQLTNNAEEIIKRIPTEAPKAKWGSTTNGLTPEQQKEYYLSKVGETFTMKAFMEA
DDILSQVNRNSQKIIVHVTDGVPTRSYAINNFKLGASYESQFEQMKKNGYLNKSNFLLTDKPEDIKGNGESYFLF
PLDSYQTQIISGNLQKLHYLDLNLNYPKGTIYRNGPVKEHGTPTKLYINSLKQKNYDIFNFGIDISGFRQVYNEE
YKKNQDGTFQKLKEEAFKLSDGEITELMRSFSSKPEYYTPIVTSADTSNNEILSKIQQQFETILTKENSIVNGTI
EDPMGDKINLQLGNGQTLQPSDYTLQGNDGSVMKDGIATGGPNNDGGILKGVKLEYIGNKLYVRGLNLGEGQKVT
LTYDVKLDDSFISNKFYDTNGRTTLNPKSEDPNTLRDFPIPKIRDVREYPTITIKNEKKLGEIEFIKVDKDNNKL
LLKGATFELQEFNEDYKLYLPIKNNNSKVVTGENGKISYKDLKDGKYQLIEAVSPEDYQKITNKPILTFEVVKGS
IKNIIAVNKQISEYHEEGDKHLITNTHIPPKGIIPMTGGKGILSFILIGGAMMSIAGGIYIWKRYKKSSDMSIKK
D

Two E boxes containing conserved glutamic residues have also been identified in GBS 67. The E box motifs are underlined in SEQ ID NO: 16 below. The conserved glutamic acid (E) residues, at amino acid residues 96 and 801, are marked in bold. The E box motifs, in particular the conserved glutamic acid residues, are thought to be important for the formation of oligomeric pilus-like structures of GBS 67. Preferred fragments of GBS 67 include at least one conserved glutamic acid residue. Preferably, fragments include at least one E box motif.

SEQ ID NO: 16

MRKYQKFSKILTLSLFCLSQIPLNTNVLGESTVPENGAKGKLVVKKTDDQNKPLSKATFVLKTTAHPESKIEKVT
AELTGEATFDNLIPGDYTLSEETAPEGYKKTNQTWQVKVESNGKTTIQNSGDKNSTIGQNQEELDKQYPPTGIYE
DTKESYKLEHVKGSVPNGKSEAKAVNPYSSEGEHIREIPEGTLSKRISEVGDLAHNKYKIELTVSGKTIVKPVDK
QKPLDVVFVLDNSNSMNNDGPNFQRHNKAKKAAEALGTAVKDILGANSDNRVALVTYGSDIFDGRSVDVVKGFKE
DDKYYGLQTKFTIQTENYSHKQLTNNAEEIIKRIPTEAPKAKWGSTTNGLTPEQQKEYYLSKVGETFTMKAFMEA
DDILSQVNRNSQKIIVHVTDGVPTRSYAINNFKLGASYESQFEQMKKNGYLNKSNFLLTDKPEDIKGNGESYFLF
PLDSYQTQIISGNLQKLHYLDLNLNYPKGTIYRNGPVKEHGTPTKLYINSLKQKNYDIFNFGIDISGFRQVYNEE
YKKNQDGTFQKLKEEAFKLSDGEITELMRSFSSKPEYYTPIVTSADTSNNEILSKIQQQFETILTKENSIVNGTI
EDPMGDKINLQLGNGQTLQPSDYTLQGNDGSVMKDGIATGGPNNDGGILKGVKLEYIGNKLYVRGLNLGEGQKVT
LTYDVKLDDSFISNKFYDTNGRTTLNPKSEDPNTLRDFPIPKIRDVREYPTITIKNEKKLGEIEFIKVDKDNNKL
LLKGATFELQEFNEDYKLYLPIKNNNSKVVTGENGKISYKDLKDGKYQLIEAVSPEDYQKITNKPILTFEVVKGS
IKNIIAVNKQISEYHEEGDKHLITNTHIPPKGIIPMTGGKGILSFILIGGAMMSIAGGIYIWKRYKKSSDMSIKK
D

Predicted secondary structure for the GBS 067 amino acid sequence is set forth in FIGURE 33. As shown in this figure, GBS 067 contains several regions predicted to form alpha helical structures. Such alpha helical regions are likely to form coiled-coil structures and may be involved in oligomerization of GBS 067.

The amino acid sequence for GBS 067 also contains a region which is homologous to the Cna_B domain of the Staphylococcus aureus collagen-binding surface protein (pfam05738). Although the Cna_B region is not thought to mediate collagen binding, it is predicted to form a beta sandwich structure. In the Staph aureus protein, this beta sandwich structure is through to form a stalk that presents the ligand binding domain away from the bacterial cell surface. This same amino acid sequence region is also predicted to be an outer membrane protein involved in cell envelope biogenesis.

The amino acid sequence for GBS 067 contains a region which is homologous to a von Willebrand factor (vWF) type A domain. The vWF type A domain is present at amino acid residues 229-402 of GBS 067 as shown in SEQ ID NO: 16. This type of sequence is typically found in extracellular proteins such as integrins and it thought to mediate adhesion, including adhesion to collagen, fibronectin, and fibrinogen, discussed above.

Because applicants have identified GBS 67 as a surface exposed protein on GBS and because GBS 67 may be involved in GBS adhesion, the immunogenicity of the GBS 67 protein was examined in mice. The results of an immunization assay with GBS 67 are set forth in Table 48, below.

Table 48: GBS 67 Protects Mice in an Immunization Assay

Challenge GBS strain (serotype)	GBS 67 immungen		PBS immunogen		FACS
	dead/treated	% survival	dead/treated	% survival	Δmean
3050 (II)	0/30	100	29/49	41	460
CJB111 (V)	76/185	59	143/189	24	481
7357 b (Ib)	34/56	39	65/74	12	316

As shown in Table 48, immunization with GBS 67 provides a substantially improved survival rate for challenged mice relative to negative control, PBS, immunized mice. These results indicate that GBS 67 may comprise an immunogenic composition of the invention.

<u>GBS 59</u>

The following offers examples of GBS 59 fragments. Nucleotide and amino acid sequences of GBS 59 sequenced from serotype V isolated strain 2603 are set forth below as SEQ ID NOS: 125 and 126. The GBS 59 polypeptide of SEQ ID NO: 126 is referred to as SAG1407.

SEQ ID NO: 125

ttaagcttcctttgattggcgtcttttcatgataactactgctccaagcataatgcttaaaccaataattgtgaa
aagaattgtaccaataccacctgtttgtgggattgttacctttttatttctacacgtgtcgcatctttttggtt
gctgttagcaacgtagtcaatgttaccacctgttatgtatgacccttgattaactacaaacttaatattacctgc
caacttagcaaatcctgctggagcaagtgtttcttcaaggttgtaagtaccgtctgcaagacctgtaacttcaaa
ttgaccttgatcgtttgaagtgtaggtaatggctctagccttatctgttatccactcataagctgtacgagcctc
aatgaaggctgcatcgtaatctgcttgtttagttttgataagttcttttgcagtaattcctttttcaccttttttg
gtctgttgcagacaacttgttataagcagcgatagcttcatctaaagctattttcttagcagctaagttttttg
accttctgattgatctgctttaagagcaaggtatttacctgctgagtttttcacaacgaattgtgcaccagccaa
acggtcaccttgttcattagttttgacaaatttcttaccatgagtttcaacttttggttcagttgggttcaatgg
tgttgggttatcagaatctttggtattggtaatggttactttaccatttctagatttattgcacttccgtaacc
agaaacacgttctgagatcatgtatgatttgttttctagaccagtgaatttacccgagaagttaccagatacttc
aaatttgataccatttccaaggtcgattgtacctttagatgtttttgtcaatgatactgaagcaacagttttatc
tttatctttcaatgtgtaaacaacgtttacaccatcaggtgcaattccgtcagaccaagttttagcaactgttac
ttcaccctttgaaggtgtaacaggaagttcagtcaagtctttacctggtttgttaccatacgacaatttgatatc
attggattctggattatcaataattgcttgaccattaacagtagcactataagtcaatgtaaattcaatatcagc
tgttttagctgctttttccaatttgcccaatccatcagctgtgaattttaatgtgaaaccacgggcatcaatgct
aagttcatagtctgtatccttagcaaaagtttctgtagttcctgaagctttaaggctaacagttgaacccattgt
caaaccatttgacattatatctgtccaaaccaagttttcgtatttagaacctttgtgaatttttgttttaacttc
ataaggaacaactttaccgatttcagcagtagcagttgctttgtcacgtgcataattaccataatttgcgccagc
tgtcaaaagtctattaacatctgtcaatgctgtcaaatcgtttgttttagcaaagtttttatcaatttctggttt
ttcttcagtgttctttggataaacatgggcatcagcaacaacaccatcttcatttaccaatggaagagtgatgtt
aactggaaccgcttttgaagcagccaggagggaaccattattgttgtaagtagattttgatttaacttcaacaat
tttaaactcgcctttcaatcctttggtgttgaaaacaagtccagtatctccctctggtgtcaatccagacacggc
ctcatcaatatttactgttatttcaggagtaccatctttattaattaaggctggtgttaatttgttaccttcttt
tgccttaacatattgcactttaccacttttatcttctttcaaagctaaagcaaagaacgcaccttcgatttcttt
agatccctcgccaaagtaaccagcaaggtcagaaatagctccacctttgtagtcttttccgttaagacctgtagt
tcctgggaagttacttttgttaagatttgattcggtttgcaaaatcttgtgcaaagtcactgtattagttgttgc

ttgatgcgcaaaggctggtgcaactgagagcaatgacgttaaagtcagtaacaatgccgagaacattgcaaaata
tttgttgattcttttcat

SEQ ID NO: 126

MKRINKYFAMFSALLLTLTSLLSVAPAFADEATTNTVTLHKILQTESNLNKSNFPGTTGLNGKDYKGGAISDLAG
YFGEGSKEIEGAFFALALKEDKSGKVQYVKAKEGNKLTPALINKDGTPEITVNIDEAVSGLTPEGDTGLVFNTKG
LKGEFKIVEVKSKSTYNNNGSLLAASKAVPVNITLPLVNEDGVVADAHVYPKNTEEKPEIDKNFAKTNDLTALTD
VNRLLTAGANYGNYARDKATATAEIGKVVPYEVKTKIHKGSKYENLVWTDIMSNGLTMGSTVSLKASGTTETFAK
DTDYELSIDARGFTLKFTADGLGKLEKAAKTADIEFTLTYSATVNGQAIIDNPESNDIKLSYGNKPGKDLTELPV
TPSKGEVTVAKTWSDGIAPDGVNVVYTLKDKDKTVASVSLTKTSKGTIDLGNGIKFEVSGNFSGKFTGLENKSYM
ISERVSGYGSAINLENGKVTITNTKDSDNPTPLNPTEPKVETHGKKFVKTNEQGDRLAGAQFVVKNSAGKYLALK
ADQSEGQKTLAAKKIALDEAIAAYNKLSATDQKGEKGITAKELIKTKQADYDAAFIEARTAYEWITDKARAITYT
SNDQGQFEVTGLADGTYNLEETLAPAGFAKLAGNIKFVVNQGSYITGGNIDYVANSNQKDATRVENKKVT*IPQTG*
GIGTILFTIIGLSIMLGAVVIMKRRQSKEA

Nucleotide and amino acid sequences of GBS 59 sequenced from serotype V isolated strain CJB111 are set forth below as SEQ ID NOS: 127 and 128. The GBS 59 polypeptide of SEQ ID NO: 128 is referred to as BO1575.

SEQ ID NO: 127

ATGAAAAAAATCAACAAATGTCTTACAATGTTCTCGACACTGCTATTGATCTTAACGTCACTATTCTCAGTTGCA
CCAGCGTTTGCGGACGACGCAACAACTGATACTGTGACCTTGCACAAGATTGTCATGCCACAAGCTGCATTTGAT
AACTTTACTGAAGGTACAAAAGGTAAGAATGATAGCGATTATGTTGGTAAACAAATTAATGACCTTAAATCTTAT
TTTGGCTCAACCGATGCTAAAGAAATCAAGGGTGCTTTCTTTGTTTTCAAAAATGAAACTGGTACAAAATTCATT
ACTGAAAATGGTAAGGAAGTCGATACTTTGGAAGCTAAAGATGCTGAAGGTGGTGCTGTTCTTTCAGGGTTAACA
AAAGACAATGGTTTTGTTTTTAACACTGCTAAGTTAAAAGGAATTTACCAAATCGTTGAATTGAAAGAAAAATCA
AACTACGATAACAACGGTTCTATCTTGGCTGATTCAAAAGCAGTTCCAGTTAAAATCACTCTGCCATTGGTAAAC
AACCAAGGTGTTGTTAAAGATGCTCACATTTATCCAAAGAATACTGAAACAAAACCACAAGTAGATAAGAACTTT
GCAGATAAAGATCTTGATTATACTGACAACCGAAAAGACAAAGGTGTTGTCTCAGCGACAGTTGGTGACAAAAAA
GAATACATAGTTGGAACAAAAATTCTTAAAGGCTCAGACTATAAGAAACTGGTTTGGACTGATAGCATGACTAAA
GGTTTGACGTTCAACAACAACGTTAAAGTAACATTGGATGGTGAAGATTTTCCTGTTTTAAACTACAAACTCGTA
ACAGATGACCAAGGTTTCCGTCTTGCCTTGAATGCAACAGGTCTTGCAGCAGTAGCAGCAGCTGCAAAAGACAAA
GATGTTGAAATCAAGATCACTTACTCAGCTACGGTGAACGGCTCCACTACTGTTGAAATTCCAGAAACCAATGAT
GTTAAATTGGACTATGGTAATAACCCAACGGAAGAAAGTGAACCACAAGAAGGTACTCCAGCTAACCAAGAAATT
AAAGTCATTAAAGACTGGGCAGTAGATGGTACAATTACTGATGCTAATGTTGCAGTTAAAGCTATCTTTACCTTG
CAAGAAAAACAAACGGATGGTACATGGGTGAACGTTGCTTCACACGAAGCAACAAAACCATCACGCTTTGAACAT
ACTTTCACAGGTTTGGATAATGCTAAAACTTACCGCGTTGTCGAACGTGTTAGCGGCTACACTCCAGAATACGTA
TCATTTAAAAATGGTGTTGTGACTATCAAGAACAACAAAAACTCAAATGATCCAACTCCAATCAACCCATCAGAA
CCAAAAGTGGTGACTTATGGACGTAAATTTGTGAAAACAAATCAAGCTAACACTGAACGCTTGGCAGGAGCTACC
TTCCTCGTTAAGAAAGAAGGCAAATACTTGGCACGTAAAGCAGGTGCAGCAACTGCTGAAGCAAAGGCAGCTGTA
AAAACTGCTAAACTAGCATTGGATGAAGCTGTTAAAGCTTATAACGACTTGACTAAAGAAAAACAAGAAGGCCAA
GAAGGTAAAACAGCATTGGCTACTGTTGATCAAAAACAAAAAGCTTACAATGACGCTTTTGTTAAAGCTAACTAC
TCATATGAATGGGTTGCAGATAAAAAGGCTGATAATGTTGTTAAATTGATCTCTAACGCCGGTGGTCAATTTGAA
ATTACTGGTTTGGATAAAGGCACTTATGGCTTGGAAGAAACTCAAGCACCAGCAGGTTATGCGACATTGTCAGGT
GATGTAAACTTTGAAGTAACTGCCACATCATATAGCAAAGGGGCTACAACTGACATCGCATATGATAAAGGCTCT
GTAAAAAAAGATGCCCAACAAGTTCAAAACAAAAAAGTAACCATCCCACAAACAGGTGGTATTGGTACAATTCTT
TTCACAATTATTGGTTTAAGCATTATGCTTGGAGCAGTAGTTATCATGAAAAAACGTCAATCAGAGGAAGCTTAA

SEQ ID NO: 128

MKKINKCLTMFSTLLLILTSLFSVAPAFADDATTDTVTLHKIVMPQAAFDNFTEGTKGKNDSDYVGKQINDLKSY
FGSTDAKEIKGAFFVFKNETGTKFITENGKEVDTLEAKDAEGGAVLSGLTKDNGFVFNTAKLKGIYQIVELKEKS
NYDNNGSILADSKAVPVKITLPLVNNQGVVKDAHIYPKNTETKPQVDKNFADKDLDYTDNRKDKGVVSATVGDKK
EYIVGTKILKGSDYKKLVWTDSMTKGLTFNNNVKVTLDGEDFPVLNYKLVTDDQGFRLALNATGLAAVAAAKDK
DVEIKITYSATVNGSTTVEIPETNDVKLDYGNNPTEESEPQEGTPANQEIKVIKDWAVDGTITDANVAVKAIFTL
QEKQTDGTWVNVASHEATKPSRFEHTFTGLDNAKTYRVVERVSGYTPEYVSFKNGVVTIKNNKNSNDPTPINPSE
PKVVTYGRKFVKTNQANTERLAGATFLVKKEGKYLARKAGAATAEAKAAVKTAKLALDEAVKAYNDLTKEKQEGQ
EGKTALATVDQKQKAYNDAFVKANYSYEWVADKKADNVVKLISNAGGQFEITGLDKGTYGLEETQAPAGYATLSG
DVNFEVTATSYSKGATTDIAYDKGSVKKDAQQVQNKKVT*IPQTG*GIGTILFTIIGLSIMLGAVVIMKKRQSEEA

The GBS 59 polypeptides contain an amino acid motif indicative of a cell wall anchor: **SEQ ID NO: 129** IPQTG (shown in italics in SEQ ID NOs: 126 and 128 above). In some recombinant host cell systems, it may be preferable to remove this motif to facilitate secretion of a recombinant GBS 59 protein from the host cell. Alternatively, in some recombinant host cell systems, it may be preferable to use the cell wall anchor motif to anchor the recombinantly expressed protein to the cell wall. The extracellular domain of the expressed protein may be cleaved during purification or the recombinant protein may be left attached to either inactivated host cells or cell membranes in the final composition.

Pilin motifs, containing conserved lysine (K) residues have been identified in the GBS 59 polypeptides. The pilin motif sequences are underlined in each of SEQ ID NOs: 126 and 128, below. Conserved lysine (K) residues are marked in bold. The conserved lysine (K) residues are located at amino acid residues 202 and 212 and amino acid residues 489 and 495 of SEQ ID NO: 126 and at amino acid residues 188 and 198 of SEQ ID NO: 128. The pilin sequences, in particular the conserved lysine residues, are thought to be important for the formation of oligomeric, pilus-like structures of GBS 59. Preferred fragments of GBS 59 include at least one conserved lysine residue. Preferably, fragments include at least one pilin sequence.

SEQ ID NO: 126

MKRINKYFAMFSALLLTLTSLLSVAPAFADEATTNTVTLHKILQTESNLNKSNFPGTTGLNGKDYKGGAISDLAG
YFGEGSKEIEGAFFALALKEDKSGKVQYVKAKEGNKLTPALINKDGTPEITVNIDEAVSGLTPEGDTGLVFNTKG
LKGEFKIVEVKSKSTYNNNGSLLAASKAVPVNITLPLVNEDGVVADAHVYP**K**NTEEKPEID**K**NFAKTNDLTALTD
VNRLLTAGANYGNYARDKATATAEIGKVVPYEVKTKIHKGSKYENLVWTDIMSNGLTMGSTVSLKASGTTETFAK
DTDYELSIDARGFTLKFTADGLGKLEKAAKTADIEFTLTYSATVNGQAIIDNPESNDIKLSYGNKPGKDLTELPV
TPSKGEVTVAKTWSDGIAPDGVNVVYTLKDKDKTVASVSLTKTSKGTIDLGNGIKFEVSGNFSGKFTGLENKSYM
ISERVSGYGSAINLENGKVTITNTKDSDNPTPLNPTEP**K**VETHG**K**KFVKTNEQGDRLAGAQFVVKNSAGKYLALK
ADQSEGQKTLAAKKIALDEAIAAYNKLSATDQKGEKGITAKELIKTKQADYDAAFIEARTAYEWITDKARAITYT
SNDQGQFEVTGLADGTYNLEETLAPAGFAKLAGNIKFVVNQGSYITGGNIDYVANSNQKDATRVENKKVTIPQTG
GIGTILFTIIGLSIMLGAVVIMKRRQSKEA

SEQ ID NO: 128

MKKINKCLTMFSTLLLILTSLFSVAPAFADDATTDTVTLHKIVMPQAAFDNFTEGTKGKNDSDYVGKQINDLKSY
FGSTDAKEIKGAFFVFKNETGTKFITENGKEVDTLEAKDAEGGAVLSGLTKDNGFVFNTAKLKGIYQIVELKEKS
NYDNNGSILADSKAVPVKITLPLVNNQGVVKDAHIYP**K**NTETKPQVD**K**NFADKDLDYTDNRKDKGVVSATVGDKK
EYIVGTKILKGSDYKKLVWTDSMTKGLTFNNNVKVTLDGEDFPVLNYKLVTDDQGFRLALNATGLAAVAAAAKDK
DVEIKITYSATVNGSTTVEIPETNDVKLDYGNNPTEESEPQEGTPANQEIKVIKDWAVDGTITDANVAVKAIFTL
QEKQTDGTWVNVASHEATKPSRFEHTFTGLDNAKTYRVVERVSGYTPEYVSFKNGVVTIKNNKNSNDPTPINPSE
PKVVTYGRKFVKTNQANTERLAGATFLVKKEGKYLARKAGAATAEAKAAVKTAKLALDEAVKAYNDLTKEKQEGQ
EGKTALATVDQKQKAYNDAFVKANYSYEWVADKKADNVVKLISNAGGQFEITGLDKGTYGLEETQAPAGYATLSG
DVNFEVTATSYSKGATTDIAYDKGSVKKDAQQVQNKKVTIPQTGGIGTILFTIIGLSIMLGAVVIMKKRQSEEA

An E box containing a conserved glutamic residue has also been identified in each of the GBS 59 polypeptides. The E box motif is underlined in each of SEQ ID NOs: 126 and 128 below. The conserved glutamic acid (E) is marked in bold at amino acid residue 621 in SEQ ID NO: 126 and at amino acid residue 588 in SEQ ID NO: 128. The E box motif, in particular the conserved glutamic acid residue, is thought to be important for the formation of oligomeric pilus-like structures of GBS 59. Preferred fragments of GBS 59 include the conserved glutamic acid residue. Preferably, fragments include the E box motif.

SEQ ID NO: 126

MKRINKYFAMFSALLLTLTSLLSVAPAFADEATTNTVTLHKILQTESNLNKSNFPGTTGLNGKDYKGGAISDLAG
YFGEGSKEIEGAFFALALKEDKSGKVQYVKAKEGNKLTPALINKDGTPEITVNIDEAVSGLTPEGDTGLVFNTKG
LKGEFKIVEVKSKSTYNNNGSLLAASKAVPVNITLPLVNEDGVVADAHVYPKNTEEKPEIDKNFAKTNDLTALTD
VNRLLTAGANYGNYARDKATATAEIGKVVPYEVKTKIHKGSKYENLVWTDIMSNGLTMGSTVSLKASGTTETFAK
DTDYELSIDARGFTLKFTADGLGKLEKAAKTADIEFTLTYSATVNGQAIIDNPESNDIKLSYGNKPGKDLTELPV
TPSKGEVTVAKTWSDGIAPDGVNVVYTLKDKDKTVASVSLTKTSKGTIDLGNGIKFEVSGNFSGKFTGLENKSYM
ISERVSGYGSAINLENGKVTITNTKDSDNPTPLNPTEPKVETHGKKFVKTNEQGDRLAGAQFVVKNSAGKYLALK
ADQSEGQKTLAAKKIALDEAIAAYNKLSATDQKGEKGITAKELIKTKQADYDAAFIEARTAYEWITDKARAITYT
SNDQGQFEVTGLADGTYNLEETLAPAGFAKLAGNIKFVVNQGSYITGGNIDYVANSNQKDATRVENKKVTIPQTG
GIGTILFTIIGLSIMLGAVVIMKRRQSKEA

SEQ ID NO: 128

MKKINKCLTMFSTLLLILTSLFSVAPAFADDATTDTVTLHKIVMPQAAFDNFTEGTKGKNDSDYVGKQINDLKSY
FGSTDAKEIKGAFFVFKNETGTKFITENGKEVDTLEAKDAEGGAVLSGLTKDNGFVFNTAKLKGIYQIVELKEKS
NYDNNGSILADSKAVPVKITLPLVNNQGVVKDAHIYPKNTETKPQVDKNFADKDLDYTDNRKDKGVVSATVGDKK
EYIVGTKILKGSDYKKLVWTDSMTKGLTFNNNVKVTLDGEDFPVLNYKLVTDDQGFRLALNATGLAAVAAAAKDK
DVEIKITYSATVNGSTTVEIPETNDVKLDYGNNPTEESEPQEGTPANQEIKVIKDWAVDGTITDANVAVKAIFTL
QEKQTDGTWVNVASHEATKPSRFEHTFTGLDNAKTYRVVERVSGYTPEYVSFKNGVVTIKNNKNSNDPTPINPSE
PKVVTYGRKFVKTNQANTERLAGATFLVKKEGKYLARKAGAATAEAKAAVKTAKLALDEAVKAYNDLTKEKQEGQ
EGKTALATVDQKQKAYNDAFVKANYSYEWVADKKADNVVKLISNAGGQFEITGLDKGTYGLEETQAPAGYATLSG
DVNFEVTATSYSKGATTDIAYDKGSVKKDAQQVQNKKVTIPQTGGIGTILFTIIGLSIMLGAVVIMKKRQSEEA

Female mice were immunized with either SAG1407 (SEQ ID NO: 126) or BO1575 (SEQ ID NO: 128) in an active maternal immunization assay. Pups bred from the immunized female mice survived GBS challenge better than control (PBS) treated mice. Results of the active maternal immunization assay using the GBS 59 immunogenic compositions are shown in Table 17, below.

TABLE 17: Active maternal immunization assay for GBS 59

Challenge GBS strain (serotype)	GBS 59		PBS		
	Dead/treated	Survival (%)	Dead/treated	Survival (%)	FACS
CJB111 (V)*	7/20	65	41/49	16	493
18RS21 (II)**	18/30	40	39/40	2.5	380

* immunized with BO1575

**immunized with SAG1407

Opsonophagocytosis assays also demonstrated that antibodies against BO1575 are opsonic for GBS serotype V, strain CJB111. See Figure 67.

GBS 52

Examples of polynucleotide and amino acid sequences for GBS 52 are set forth below. SEQ ID NO: 20 and 21 represent GBS 52 sequences from GBS serotype V, strain isolate 2603.

SEQ ID NO: 20

ATGAAACAAACATTAAAACTTATGTTTTCTTTTCTGTTGATGTTAGGGACTATGTTTGGAATTAGCCAAACTGTT
TTAGCGCAAGAAACTCATCAGTTGACGATTGTTCATCTTGAAGCAAGGGATATTGATCGTCCAAATCCACAGTTG
GAGATTGCCCCTAAAGAAGGGACTCCAATTGAAGGAGTACTCTATCAGTTGTACCAATTAAAATCAACTGAAGAT
GGCGATTTGTTGGCACATTGGAATTCCCTAACTATCACAGAATTGAAAAAACAGGCGCAGCAGGTTTTTGAAGCC
ACTACTAATCAACAAGGAAAGGCTACATTTAACCAACTACCAGATGGAATTTATTATGGTCTGGCGGTTAAAGCC
GGTGAAAAAAATCGTAATGTCTCAGCTTTCTTGGTTGACTTGTCTGAGGATAAAGTGATTTATCCTAAAATCATC
TGGTCCACAGGTGAGTTGGACTTGCTTAAAGTTGGTGTGGATGGTGATACCAAAAAACCACTAGCAGGCGTTGTC
TTTGAACTTTATGAAAAGAATGGTAGGACTCCTATTCGTGTGAAAAATGGGGTGCATTCTCAAGATATTGACGCT
GCAAAACATTTAGAAACAGATTCATCAGGGCATATCAGAATTTCCGGGCTCATCCATGGGGACTATGTCTTAAAA
GAAATCGAGACACAGTCAGGATATCAGATCGGACAGGCAGAGACTGCTGTGACTATTGAAAAATCAAAAACAGTA

ACAGTAACGATTGAAAATAAAAAAGTTTCCGACACCTAAAGTGCCATCTCGAGGAGGTCTTATTCCCAAAACAGGT
GAGCAACAGGCAATGGCACTTGTAATTATTGGTGGTATTTTAATTGCTTTAGCCTTACGATTACTATCAAAACAT
CGGAAACATCAAAATAAGGAT

SEQ ID NO: 21
MKQTLKLMFSFLLMLGTMFGISQTVLAQETHQLTIVHLEARDIDRPNPQLEIAPKEGTPIEGVLYQLYQLKSTED
GDLLAHWNSLTITELKKQAQQVFEATTNQQGKATFNQLPDGIYYGLAVKAGEKNRNVSAFLVDLSEDKVIYPKII
WSTGELDLLKVGVDGDTKKPLAGVVFELYEKNGRTPIRVKNGVHSQDIDAAKHLETDSSGHIRISGLIHGDYVLK
EIETQSGYQIGQAETAVTIEKSKTVTVTIENKKVPTPKVPSRGGL*IPKTG*EQQAMALVIIGGILIALALRLLSKH
RKHQNKD

GBS 52 contains an amino acid motif indicative of a cell wall anchor: **SEQ ID NO: 124** IPKTG (shown in italics in SEQ ID NO: 21, above). In some recombinant host cell systems, it may be preferable to remove this motif to facilitate secretion of a recombinant GBS 52 protein from the host cell. Alternatively, in other recombinant host cell systems, it may be preferable to use the cell wall anchor motif to anchor the recombinantly expressed protein to the cell wall. The extracellular domain of the expressed protein may be cleaved during purification or the recombinant protein may be left attached to either inactivated host cells or cell membranes in the final composition.

A pilin motif, discussed above, containing a conserved lysine (K) residue has also been identified in GBS 52. The pilin motif sequence is underlined in SEQ ID NO: 21, below. Conserved lysine (K) residues are also marked in bold, at amino acid residues 148 and 160. The pilin sequence, in particular the conserved lysine residues, are thought to be important for the formation of oligomeric, pilus-like structures. Preferred fragments of GBS 52 include at least one conserved lysine residue. Preferably, fragments include the pilin sequence.

SEQ ID NO: 21
MKQTLKLMFSFLLMLGTMFGISQTVLAQETHQLTIVHLEARDIDRPNPQLEIAPKEGTPIEGVLYQLYQLKSTED
GDLLAHWNSLTITELKKQAQQVFEATTNQQGKATFNQLPDGIYYGLAVKAGEKNRNVSAFLVDLSEDKVIYPKII
WSTGELDLLKVGVDGDTKKPLAGVVFELYEKNGRTPIRVKNGVHSQDIDAAKHLETDSSGHIRISGLIHGDYVLK
EIETQSGYQIGQAETAVTIEKSKTVTVTIENKKVPTPKVPSRGGLIPKTGEQQAMALVIIGGILIALALRLLSKH
RKHQNKD

An E box containing a conserved glutamic residue has been identified in GBS 52. The E-box motif is underlined in SEQ ID NO: 21, below. The conserved glutamic acid (E), at amino acid residue 226, is marked in bold. The E box motif, in particular the conserved glutamic acid residue, is thought to be important for the formation of oligomeric pilus-like structures of GBS 52. Preferred fragments of GBS 52 include the conserved glutamic acid residue. Preferably, fragments include the E box motif.

SEQ ID NO: 21
MKQTLKLMFSFLLMLGTMFGISQTVLAQETHQLTIVHLEARDIDRPNPQLEIAPKEGTPIEGVLYQLYQLKSTED
GDLLAHWNSLTITELKKQAQQVFEATTNQQGKATFNQLPDGIYYGLAVKAGEKNRNVSAFLVDLSEDKVIYPKII
WSTGELDLLKVGVDGDTKKPLAGVVFELYEKNGRTPIRVKNGVHSQDIDAAKHLETDSSGHIRISGLIHGDYVLK
EIETQSGYQIGQAETAVTIEKSKTVTVTIENKKVPTPKVPSRGGLIPKTGEQQAMALVIIGGILIALALRLLSKH
RKHQNKD

SAG0647

Examples of polynucleotide and amino acid sequences for SAG0647 are set forth below. SEQ ID NO: 22 and 23 represent SAG0647 sequences from GBS serotype V, strain isolate 2603.

SEQ ID NO: 22
ATGGGACAAAAATCAAAAATATCTCTAGCTACGAATATTCGTATATGGATTTTTCGTTTAATTTTCTTAGCGGGT
TTCCTTGTTTTGGCATTTCCCATCGTTAGTCAGGTCATGTACTTTCAAGCCTCTCACGCCAATATTAATGCTTTT
AAAGAAGCTGTTACCAAGATTGACCGGGTGGAGATTAATCGGCGTTTAGAACTTGCTTATGCTTATAACGCCAGT
ATAGCAGGTGCCAAAACTAATGGCGAATATCCAGCGCTTAAAGACCCCTACTCTGCTGAACAAAAGCAGGCAGGG
GTCGTTGAGTACGCCCGCATGCTTGAAGTCAAAGAACAAATAGGTCATGTGATTATTCCAAGAATTAATCAGGAT
ATCCCTATTTACGCTGGCTCTGCTGAAGAAAATCTTCAGAGGGGCGTTGGACATTTAGAGGGGACCAGTCTTCCA
GTCGGTGGTGAGTCAACTCATGCCGTTCTAACTGCCCATCGAGGGCTACCAACGGCCAAGCTATTTACCAATTTA
GACAAGGTAACAGTAGGTGACCGTTTTTACATTGAACACATCGGCGGAAAGATTGCTTATCAGGTAGACCAAATC
AAAGTTATCGCCCCTGATCAGTTAGAGGATTTGTACGTGATTCAAGGAGAAGATCACGTCACCCTATTAACTTGC
ACACCTTATATGATAAATAGTCATCGCCTCCTCGTTCGAGGCAAGCGAATTCCTTATGTGGAAAAAACAGTGCAG
AAAGATTCAAAGACCTTCAGGCAACAACAATACCTAACCTATGCTATGTGGGTAGTCGTTGGACTTATCTTGCTG
TCGCTTCTCATTTGGTTTAAAAAGACGAAACAGAAAAAGCGGAGAAAGAATGAAAAAGCGGCTAGTCAAAATAGT
CACAATAATTCGAAATAA

SEQ ID NO: 23
MGQKSKISLATNIRIWIFRLIFLAGFLVLAFPIVSQVMYFQASHANINAFKEAVTKIDRVEINRRLELAYAYNAS
IAGAKTNGEYPALKDPYSAEQKQAGVVEYARMLEVKEQIGHVIIPRINQDIPIYAGSAEENLQRGVGHLEGTSLP
VGGESTHAVLTAHRGLPTAKLFTNLDKVTVGDRFYIEHIGGKIAYQVDQIKVIAPDQLEDLYVIQGEDHVTLLTC
TPYMINSHRLLVRGKRIPYVEKTVQKDSKTFRQQQYLTYAMWVVVGLILLSLLIWFKKTKQKKRRKNEKAASQNS
HNNSK

SAG0648

Examples of polynucleotide and amino acid sequences for SAG0648 are set forth below.

SEQ ID NO: 24 and 25 represent SAG0648 sequences from GBS serotype V, strain isolate 2603.

SEQ ID NO: 24
ATGGGAAGTCTGATTCTCTTATTTCCGATTGTGAGCCAGGTAAGTTACTACCTTGCTTCGCATCAAAATATTAAT
CAATTTAAGCGGGAAGTCGCTAAGATTGATACTAATACGGTTGAACGACGCATCGCTTTAGCTAATGCTTACAAT
GAGACGTTATCAAGGAATCCCTTGCTTATAGACCCTTTTACCAGTAAGCAAAAAGAAGGTTTGAGAGAGTATGCT
CGTATGCTTGAAGTTCATGAGCAAATAGGTCATGTGGCAATCCCAAGTATTGGGGTTGATATTCCAATTTATGCT
GGAACATCCGAAACTGTGCTTCAGAAAGGTAGTGGGCATTTGGAGGGAACCAGTCTTCCAGTGGGAGGTTTGTCA
ACCCATTCAGTACTAACTGCCCACCGTGGCTTGCCAACAGCTAGGCTATTTACCGACTTAAATAAAGTTAAAAAA
GGCCAGATTTTCTATGTGACGAACATCAAGGAAACACTTGCCTACAAAGTCGTGTCTATCAAAGTTGTGGATCCA
ACAGCTTTAAGTGAGGTTAAGATTGTCAATGGTAAGGATTATATAACCTTGCTGACTTGCACACCTTACATGATC
AATAGTCATCGTCTCTTGGTAAAAGGAGAGCGTATTCCTTATGATTCTACCGAGGCGGAAAAGCACAAAGAACAA
ACCGTACAAGATTATCGTTTGTCACTAGTGTTGAAGATACTACTAGTATTATTAATTGGACTCTTCATCGTGATA
ATGATGAGAAGATGGATGCAACATCGTCAATAA

SEQ ID NO: 25
MGSLILLFPIVSQVSYYLASHQNINQFKREVAKIDTNTVERRIALANAYNETLSRNPLLIDPFTSKQKEGLREYA
RMLEVHEQIGHVAIPSIGVDIPIYAGTSETVLQKGSGHLEGTSLPVGGLSTHSVLTAHRGLPTARLFTDLNKVKK
GQIFYVTNIKETLAYKVVSIKVVDPTALSEVKIVNGKDYITLLTCTPYMINSHRLLVKGERIPYDSTEAEKHKEQ
TVQDYRLSLVLKILLVLLIGLFIVIMMRRWMQHRQ

GBS 150

Examples of polynucleotide and amino acid sequences for GBS 150 are set forth below. SEQ ID NO: 26 and 27 represent GBS 150 sequences from GBS serotype V, strain isolate 2603.

SEQ ID NO: 26
ATGAAAAAGATTAGAAAAAGTTTAGGACTTCTACTATGTTGCTTTTTAGGATTGGTACAATTAGCGTTTTTTTCG
GTAGCCAGTGTAAATGCTGATACCCCTAATCAACTAACAATCACACAGATAGGACTTCAGCCAAATACTACAGAG
GAGGGGATTTCTTATCGTTTATGGACTGTGACTGACAACTTAAAAGTTGATTTATTGAGCCAAATGACAGATAGC
GAATTGAACCAGAAGTATAAGAGTATCTTGACTTCTCCTACTGATACTAATGGTCAGACAAAGATAGCACTCCCA
AATGGTTCGTACTTTGGTCGTGCTTATAAAGCTGATCAAAGCGTTTCAACAATAGTACCTTTTTATATTGAATTA
CCAGATGATAAGTTATCAAATCAATTACAGATAAATCCTAAGCGAAAAGTTGAAACAGGCCGATTAAAACTTATT

AAATATACAAAAGAAGGAAAGATAAAGAAAAGGCTATCCGGAGTAATATTTGTATTATACGATAACCAGAATCAG
CCAGTTCGCTTTAAAAATGGACGATTTACGACCGATCAAGATGGGATTACTTCATTAGTAACTGATGATAAGGGA
GAAATTGAGGTTGAAGGTTTATTACCTGGTAAGTATATTTTTCGAGAAGCAAAAGCACTAACTGGTTACCGTATA
TCTATGAAGGATGCTGTAGTTGCTGTAGTTGCTAATAAAACACAGGAAGTAGAGGTAGAAAACGAAAAAGAAACT
CCTCCACCAACAAATCCTAAACCATCACAACCGCTTTTTCCACAATCATTTCTTCCTAAAACAGGAATGATTATT
GGTGGAGGACTGACAATTCTTGGTTGTATTATTTTGGGAATTTTGTTTATCTTTTTAAGAAAAACTAAAAATAGC
AAATCTGAAAGAAACGATACAGTA

SEQ ID NO: 27
MKKIRKSLGLLLCCFLGLVQLAFFSVASVNADTPNQLTITQIGLQPNTTEEGISYRLWTVTDNLKVDLLSQMTDS
ELNQKYKSILTSPTDTNGQTKIALPNGSYFGRAYKADQSVSTIVPFYIELPDDKLSNQLQINPKRKVETGRLKLI
KYTKEGKIKKRLSGVIFVLYDNQNQPVRFKNGRFTTDQDGITSLVTDDKGEIEVEGLLPGKYIFREAKALTGYRI
SMKDAVVAVVANKTQEVEVENEKETPPPTNPKPSQPLFPQSF*LPKTG*MIIGGGLTILGCIILGILFIFLRKTKNS
KSERNDTV

GBS 150 contains an amino acid motif indicative of a cell wall anchor: **SEQ ID NO: 130** LPKTG (shown in italics in SEQ ID NO: 27 above). In some recombinant host cell systems, it may be preferable to remove this motif to facilitate secretion of a recombinant GBS 150 protein from the host cell. Alternatively, in other recombinant host cell systems, it may be preferable to use the cell wall anchor motif to anchor the recombinantly expressed protein to the cell wall. The extracellular domain of the expressed protein may be cleaved during purification or the recombinant protein may be left attached to either inactivated host cells or cell membranes in the final composition.

As discussed above, a pilin motif, containing a conserved lysine (K) residue has been identified in GBS 150. The pilin motif sequence is underlined in SEQ ID NO: 27, below. Conserved lysine (K) residues are marked in bold, at amino acid residues 139 and 148. The pilin sequence, in particular the conserved lysine residues, are thought to be important for the formation of oligomeric, pilus-like structures of GBS 150. Preferred fragments of GBS 150 include a conserved lysine residue. Preferably, fragments include the pilin sequence.

SEQ ID NO: 27
MKKIRKSLGLLLCCFLGLVQLAFFSVASVNADTPNQLTITQIGLQPNTTEEGISYRLWTVTDNLKVDLLSQMTDS
ELNQKYKSILTSPTDTNGQTKIALPNGSYFGRAYKADQSVSTIVPFYIELPDDKLSNQLQINP**K**RKVETGRL**K**LI
KYTKEGKIKKRLSGVIFVLYDNQNQPVRFKNGRFTTDQDGITSLVTDDKGEIEVEGLLPGKYIFREAKALTGYRI
SMKDAVVAVVANKTQEVEVENEKETPPPTNPKPSQPLFPQSFLPKTGMIIGGGLTILGCIILGILFIFLRKTKNS
KSERNDTV

An E box containing a conserved glutamic residue has also been identified in GBS 150. The E box motif is underlined in SEQ ID NO: 27 below. The conserved glutamic acid (E), at amino acid residue 216, is marked in bold. The E box motif, in particular the conserved glutamic acid residue, is thought to be important for the formation of oligomeric pilus-like structures of GBS 150. Preferred fragments of GBS 150 include the conserved glutamic acid residue. Preferably, fragments include the E box motif.

SEQ ID NO: 27
MKKIRKSLGLLLCCFLGLVQLAFFSVASVNADTPNQLTITQIGLQPNTTEEGISYRLWTVTDNLKVDLLSQMTDS
ELNQKYKSILTSPTDTNGQTKIALPNGSYFGRAYKADQSVSTIVPFYIELPDDKLSNQLQINPKRKVETGRLKLI
KYTKEGKIKKRLSGVIFVLYDNQNQPVRFKNGRFTTDQDGITSLVTDDKGEIEVEGLLPGKYIFR**E**AKALTGYRI
SMKDAVVAVVANKTQEVEVENEKETPPPTNPKPSQPLFPQSFLPKTGMIIGGGLTILGCIILGILFIFLRKTKNS
KSERNDTV

SAG1405

Examples of polynucleotide and amino acid sequences for SAG1405 are set forth below. SEQ ID NO: 28 and 29 represent SAG1405 sequences from GBS serotype V, strain isolate 2603.

SEQ ID NO: 28
ATGGGAGGAAAATTTCAGAAAAACCTTAAGAAATCGGTCGTTTTAAATCGATGGATGAATGTAGGCTTGATACTA
TTGTTCTTAGTTGGTCTTTTGATAACCTCATATCCTTTTATTTCAAATTGGTACTATAATATTAAAGCTAATAAT
CAAGTAACTAACTTTGATAATCAAACCCAAAAATTAAATACTAAAGAGATTAATAGACGATTTGAGTTAGCAAAA
GCTTATAATAGAACACTGGACCCAAGCCGCCTATCAGATCCCTATACTGAAAAAGAAAAAAAAGGTATTGCTGAA
TACGCCCACATGCTTGAGATTGCTGAAATGATTGGATATATTGATATACCGTCTATCAAGCAAAAATTACCTATC
TATGCGGGGACTACCAGTAGTGTTCTTGAAAAAGGAGCAGGACACCTTGAAGGAACCTCCTTGCCAATTGGTGGA
AAAAGTTCACATACTGTTATCACAGCTCATCGCGGCTTACCTAAAGCTAAGTTATTTACAGATTTAGATAAACTT
AAAAAAGGAAAAATTTTTTATATTCATAATATCAAAGAAGTTTTAGCCTATAAGGTTGATCAAATAAGTGTTGTA
AAGCCAGATAATTTTTCTAAATTATTGGTTGTTAAAGGTAAGGATTATGCGACTTTGCTAACATGTACACCTTAT
TCGATTAATTCACATCGTTTACTAGTTAGAGGGCATCGAATCAAGTATGTACCTCCTGTTAAAGAAAAGAACTAT
TTAATGAAAGAATTGCAAACACACTATAAACTTTATTTCCTCTTATCAATCCTAGTTATTCTTATATTAGTCGCT
TTACTATTATATTTAAAACGAAAATTTAAAGAGAGAAAGAGAAAGGGAAATCAAAAATGA

SEQ ID NO: 29
MGGKFQKNLKKSVVLNRWMNVGLILLFLVGLLITSYPFISNWYYNIKANNQVTNFDNQTQKLNTKEINRRFELAK
AYNRTLDPSRLSDPYTEKEKKGIAEYAHMLEIAEMIGYIDIPSIKQKLPIYAGTTSSVLEKGAGHLEGTSLPIGG
KSSHTVITAHRGLPKAKLFTDLDKLKKGKIFYIHNIKEVLAYKVDQISVVKPDNFSKLLVVKGKDYATLLTCTPY
SINSHRLLVRGHRIKYVPPVKEKNYLMKELQTHYKLYFLLSILVILILVALLLYLKRKFKERKRKGNQK

SAG1406

Examples of polynucleotide and amino acid sequences for SAG1405 are set forth below. SEQ ID NO: 30 and 31 represent SAG1405 sequences from GBS serotype V, strain isolate 2603.

SEQ ID NO: 30
GTGAAGACTAAAAAAATCATCAAAAAAACAAAAAAAAAGAAGAAGTCAAATCTTCCTTTTATCATTCTTTTTCTA
ATAGGTCTATCTATTTTATTGTATCCAGTGGTATCACGTTTTTACTATACGATAGAATCTAATAATCAAACACAG
GATTTTGAGAGAGCTGCTAAAAAACTTAGTCAGAAAGAAATCAATCGACGTATGGCTCTAGCACAAGCTTATAAT
GATTCTTTAAATAATGTCCATCTTGAAGATCCTTATGAGAAAAAACGAATTCAAAAGGGGGTAGCAGAGTACGCC
CGTATGTTAGAGGTAAGTGAAAAAATCGGAACAATTTCAGTTCCTAAGATAGGTCAAAAACTCCCTATATTTGCA
GGTTCAAGTCAAGAAGTTCTATCTAAAGGAGCAGGGCATTTAGAAGGTACCTCTCTTCCAATTGGGGGCAATAGT
ACACATACTGTTATAACAGCGCATTCAGGAATTCCAGATAAAGAACTCTTTTCTAACCTTAAAAAGTTAAAAAAA
GGAGATAAGTTTTATATTCAAAACATAAAAGAAACGATAGCATATCAAGTAGATCAGATAAAAGTCGTTACACCC
GATAACTTTTCAGATTTGTTGGTTGTTCCTGGACATGATTATGCAACCTTATTGACTTGCACCCCGATTATGATC
AATACACACAGACTTTTAGTAAGGGGACATCGTATCCCTTATAAAGGTCCTATTGATGAAAAATTAATAAAAGAC
GGTCATTTAAACACGATTTATAGATATCTATTCTATATATCTTTAGTTATTATTGCTTGGTTACTTTGGTTAATA
AAACGTCAACGTCAAAAAAATCGTTTAGCAAGTGTTAGAAAAGGAATTGAATCATAA

SEQ ID NO: 31
MKTKKIIKKTKKKKKSNLPFIILFLIGLSILLYPVVSRFYYTIESNNQTQDFERAAKKLSQKEINRRMALAQAYN
DSLNNVHLEDPYEKKRIQKGVAEYARMLEVSEKIGTISVPKIGQKLPIFAGSSQEVLSKGAGHLEGTSLPIGGNS
THTVITAHSGIPDKELFSNLKKLKKGDKFYIQNIKETIAYQVDQIKVVTPDNFSDLLVVPGHDYATLLTCTPIMI
NTHRLLVRGHRIPYKGPIDEKLIKDGHLNTIYRYLFYISLVIIAWLLWLIKRQRQKNRLASVRKGIES

01520

An example of an amino acid sequence for 01520 is set forth below. SEQ ID NO: 32 represents a 01520 sequence from GBS serotype III, strain isolate COH1.

SEQ ID NO: 32
MIRRYSANFLAILGIILVSSGIYWGWYNINQAHQADLTSQHIVKVLDKSITHQVKGSENGELPVKKLDKTDYLGT
LDIPNLKLHLPVAANYSFEQLSKTPTRYYGSYLTNNMVICAHNFPYHFDALKNVDMGTDVYFTTTTGQIYHYKIS
NREIIEPTAIEKVYKTATSDNDWDLSLFTCTKAGVARVLVRCQLIDVKN

01521

An example of an amino acid sequence for 01521 is set forth below. SEQ ID NO: 33 represents a 01521 sequence from GBS serotype III, strain isolate COH1.

SEQ ID NO: 33
MIYKKILKITLLLLFSLSTQLVSADTNDQMKTGSITIQNKYNNQGIAGGNLLVYQVAQAKDVDGNQVFTLTTPFQ
GIGIKDDDLTQVNLDSNQAKYVNLLTKAVHKTQPLQTFDNLPAEGIVANNLPQGIYLFIQTKTAQGYELMSPFIL
SIPKDGKYDITAFEKMSPLNAKPKKEETITPTVTHQTKGK*LPFTG*QVWWPIPILIMSGLLCLIIALKWRRRRD

01521 contains an amino acid motif indicative of a cell wall anchor: **SEQ ID NO: 132** LPFTG (shown in italics in SEQ ID NO: 33 above). In some recombinant host cell systems, it may be preferable to remove this motif to facilitate secretion of a recombinant 01521 protein from the host cell. Alternatively, it may be preferable to use the cell wall anchor motif to anchor the recombinantly expressed protein to the cell wall. The extracellular domain of the expressed protein may be cleaved during purification or the recombinant protein may be left attached to either inactivated host cells or cell membranes in the final composition.

Two pilin motifs, containing conserved lysine (K) residues have been identified in 01521. The pilin motif sequences are underlined in SEQ ID NO: 33, below. Conserved lysine (K) residues are marked in bold, at amino acid residues 154 and 165 and at amino acid residues 174 and 188. The pilin sequences, in particular the conserved lysine residues, are thought to be important for the formation of oligomeric, pilus-like structures of 01521. Preferred fragments of 01521 include at least one conserved lysine residue. Preferably, fragments include at least one pilin sequence.

SEQ ID NO: 33
MIYKKILKITLLLLFSLSTQLVSADTNDQMKTGSITIQNKYNNQGIAGGNLLVYQVAQAKDVDGNQVFTLTTPFQ
GIGIKDDDLTQVNLDSNQAKYVNLLTKAVHKTQPLQTFDNLPAEGIVANNLPQGIYLFIQTKTAQGYELMSPFIL
SIP**K**DGKYDITAFE**K**MSPLNAKP**K**KEETITPTVTHQT**K**GKLPFTGQVWWPIPILIMSGLLCLIIALKWRRRRD

An E box containing a conserved glutamic residue has also been identified in 01521. The E box motif is underlined in SEQ ID NO: 33 below. The conserved glutamic acid (E), at amino acid residue 177, is marked in bold. The E box motif, in particular the conserved glutamic acid residue, is thought to be important for the formation of oligomeric pilus-like structures of 01521. Preferred fragments of 01521 include the conserved glutamic acid residue. Preferably, fragments include the E box motif.

SEQ ID NO: 33
MIYKKILKITLLLLFSLSTQLVSADTNDQMKTGSITIQNKYNNQGIAGGNLLVYQVAQAKDVDGNQVFTLTTPFQ
GIGIKDDDLTQVNLDSNQAKYVNLLTKAVHKTQPLQTFDNLPAEGIVANNLPQGIYLFIQTKTAQGYELMSPFIL
SIPKDGKYDITAFEKMSPLNAKPKKE**E**TITPTVTHQTKGKLPFTGQVWWPIPILIMSGLLCLIIALKWRRRRD

01522

An example of an amino acid sequence for 01522 is set forth below. SEQ ID NO: 34 represents a 01522 sequence from GBS serotype III, strain isolate COH1.

SEQ ID NO: 34
MAYPSLANYWNSFHQSRAIMDYQDRVTHMDENDYKKIINRAKEYNKQFKTSGMKWHMTSQERLDYNSQLAIDKTG
NMGYISIPKINIKLPLYHGTSEKVLQTSIGHLEGSSLPIGGDSTHSILSGHRGLPSSRLFSDLDKLKVGDHWTVS
ILNETYTYQVDQIRTVKPDDLRDLQIVKGKDYQTLVTCTPYGVNTHRLLVRGHRVPNDNGNALVVAEAIQIEPIY
IAPFIAIFLTLILLLISLEVTRRARQRKKILKQAMRKEENNDL

01523

An example of an amino acid sequence for 01523 is set forth below. SEQ ID NO: 35 represents a 01523 sequence from GBS serotype III, strain isolate COH1.

SEQ ID NO: 35
MKKKMIQSLLVASLAFGMAVSPVTPIAFAAETGTITVQDTQKGATYKAYKVFDAEIDNANVSDSNKDGASYLIPQ
GKEAEYKASTDFNSLFTTTTNGGRTYVTKKDTASANEIATWAKSISANTTPVSTVTESNNDGTEVINVSQYGYYY
VSSTVNNGAVIMVTSVTPNATIHEKNTDATWGDGGGKTVDQKTYSVGDTVKYTITYKNAVNYHGTEKVYQYVIKD
TMPSASVVDLNEGSYEVTITDGSGNITTLTQGSEKATGKYNLLEENNNFTITIPWAATNTPTGNTQNGANDDFFY
KGINTITVTYTGVLKSGAKPGSADLPENTNIATINPNTSNDDPGQKVTVRDGQITIKKIDGSTKASLQGAIFVLK
NATGQFLNFNDTNNVEWGTEANATEYTTGADGIITITGLKEGTYYLVEKKAPLGYNLLDNSQKVILGDGATDTTN
SDNLLVNPTVENNKGTE*LPSTG*GIGTTIFYIIGAILVIGAGIVLVARRRLRS

01523 contains an amino acid motif indicative of a cell wall anchor: **SEQ ID NO: 131** LPSTG (shown in italics in SEQ ID NO: 35 above). In some recombinant host cell systems, it may be preferable to remove this motif to facilitate secretion of a recombinant 01523 protein from the host cell. Alternatively, it may be preferable to use the cell wall anchor motif to anchor the recombinantly expressed protein to the cell wall. The extracellular domain of the expressed protein may be cleaved during purification or the recombinant protein may be left attached to either inactivated host cells or cell membranes in the final composition.

An E box containing a conserved glutamic residue has also been identified in 01523. The E box motif is underlined in SEQ ID NO: 35 below. The conserved glutamic acid (E), at amino acid residue 423, is marked in bold. The E box motif, in particular the conserved glutamic acid residue, is thought to be important for the formation of oligomeric pilus-like structures of 01523. Preferred fragments of 01523 include the conserved glutamic acid residue. Preferably, fragments include the E box motif.

SEQ ID NO: 35
MKKKMIQSLLVASLAFGMAVSPVTPIAFAAETGTITVQDTQKGATYKAYKVFDAEIDNANVSDSNKDGASYLIPQ
GKEAEYKASTDFNSLFTTTTNGGRTYVTKKDTASANEIATWAKSISANTTPVSTVTESNNDGTEVINVSQYGYYY
VSSTVNNGAVIMVTSVTPNATIHEKNTDATWGDGGGKTVDQKTYSVGDTVKYTITYKNAVNYHGTEKVYQYVIKD
TMPSASVVDLNEGSYEVTITDGSGNITTLTQGSEKATGKYNLLEENNNFTITIPWAATNTPTGNTQNGANDDFFY
KGINTITVTYTGVLKSGAKPGSADLPENTNIATINPNTSNDDPGQKVTVRDGQITIKKIDGSTKASLQGAIFVLK
NATGQFLNFNDTNNVEWGTEANATEYTTGADGIITITGLKEGTYYLV**E**KKAPLGYNLLDNSQKVILGDGATDTTN
SDNLLVNPTVENNKGTELPSTGGIGTTIFYIIGAILVIGAGIVLVARRRLRS

01524

An example of an amino acid sequence for 01524 is set forth below. SEQ ID NO: 36 represents a 01524 sequence from GBS serotype III, strain isolate COH1.

SEQ ID NO: 36
MLKKCQTFIIESLKKKKHPKEWKIIMWSLMILTTFLTTYFLILPAITVEETKTDDVGITLENKNSSQVTSSTSSS
QSSVEQSKPQTPASSVTETSSSEEAAYREEPLMFRGADYTVTVTLTKEAKIPKNADLKVTELKDNSATFKDYKKK
ALTEVAKQDSEIKNFKLYDITIESNGKEAEPQAPVKVEVNYDKPLEASDENLKVVHFKDDGQTEVLKSKDTAETK
NTSSDVAFKTDSFSIYAIVQEDNTEVPRLTYHFQNNDGTDYDFLTASGMQVHHQIIKDGESLGEVGIPTIKAGEH
FNGWYTYDPTTGKYGDPVKFGEPITVTETKEICVRPFMSKVATVTLYDDSAGKSILERYQVPLDSSGNGTADLSS
FKVSPPTSTLLFVGWSKTQNGAPLSESEIQALPVSSDISLYPVFKESYGVEFNTGDLSTGVTYIAPRRVLTGQPA
STIKPNDPTRPGYTFAGWYTAASGGAAFDFNQVLTKDTTLYAHWSPAQTTYTINYWQQSATDNKNATDAQKTYEY
AGQVTRSGLSLSNQTLTQQDINDKLPTGFKVNNTRTETSVMIKDDGSSVVNVYYDRKLITIKFAKYGGYSLPEYY
YSYNWSSDADTYTGLYGTTLAANGYQWKTGAWGYLANVGNNQVGTYGMSYLGEFILPNDTVDSDVIKLFPKGNIV
QTYRFFKQGLDGTYSLADTGGGAGADEFTFTEKYLGFNVKYYQRLYPDNYLFDQYASQTSAGVKVPISDEYYDRY
GAYHKDYLNLVVWYERNSYKIKYLDPLDNTELPNFPVKDVLYEQNLSSYAPDTTTVQPKPSRPGYVWDGKWYKDQ
AQTQVFDFNTTMPPHDVKVYAGWQKVTYRVNIDPNGGRLSKTDDTYLDLHYGDRIPDYTDITRDYIQDPSGTYYY
KYDSRDKDPDSTKDAYYTTDTSLSNVDTTTKYKYVKDAYKLVGWYYVNPDGSIRPYNFSGAVTQDINLRAIWRKA

GDYHIIYSNDAVGTDGKPALDASGQQLQTSNEPTDPDSYDDGSHSALLRRPTMPDGYRFRGWWYNGKIYNPYDSI
DIDAHLADANKNITIKPVIIPVGDIKLEDTSIKYNGNGGTRVENGNVVTQVETPRMELNSTTTIPENQYFTRTGY
NLIGWHHDKDLADTGRVEFTAGQSIGIDNNPDATNTLYAVWQPKEYTVRVSKTVVGLDEDKTKDFLFNPSETLQQ
ENFPLRDGQTKEFKVPYGTSISIDEQAYDEFKVSESITEKNLATGEADKTYDATGLQSLTVSGDVDISFTNTRIK
QKVRLQKVNVENDNNFLAGAVFDIYESDANGNKASHPMYSGLVTNDKGLLLVDANNYLSLPVGKYYLTETKAPPG
YLLPKNDISVLVISTGVTFEQNGNNATPIKENLVDGSTVYTFKITNSKGTE*LPSTG*GIGTHIYILVGLALALPSG
LILYYRKKI

01524 contains an amino acid motif indicative of a cell wall anchor: **SEQ ID NO: 131** LPSTG (shown in italics in SEQ ID NO: 36 above). In some recombinant host cell systems, it may be preferable to remove this motif to facilitate secretion of a recombinant 01524 protein from the host cell. Alternatively, it may be preferable to use the cell wall anchor motif to anchor the recombinantly expressed protein to the cell wall. The extracellular domain of the expressed protein may be cleaved during purification or the recombinant protein may be left attached to either inactivated host cells or cell membranes in the final composition.

Three pilin motifs, containing conserved lysine (K) residues have been identified in 01524. The pilin motif sequences are underlined in SEQ ID NO: 36, below. Conserved lysine (K) residues are marked in bold, at amino acid residues 128 and 138, amino acid residues 671 and 682, and amino acid residues 809 and 820. The pilin sequences, in particular the conserved lysine residues, are thought to be important for the formation of oligomeric, pilus-like structures of 01524. Preferred fragments of 01524 include at least one conserved lysine residue. Preferably, fragments include at least one pilin sequence.

SEQ ID NO: 36
MLKKCQTFIIESLKKKKHPKEWKIIMWSLMILTTFLTTYFLILPAITVEETKTDDVGITLENKNSSQVTSSTSSS
QSSVEQSKPQTPASSVTETSSSEEAAYREEPLMFRGADYTVTVTLTKEAKIP**K**NADLKVTEL**K**DNSATFKDYKKK
ALTEVAKQDSEIKNFKLYDITIESNGKEAEPQAPVKVEVNYDKPLEASDENLKVVHFKDDGQTEVLKSKDTAETK
NTSSDVAFKTDSFSIYAIVQEDNTEVPRLTYHFQNNDGTDYDFLTASGMQVHHQIIKDGESLGEVGIPTIKAGEH
FNGWYTYDPTTGKYGDPVKFGEPITVTETKEICVRPFMSKVATVTLYDDSAGKSILERYQVPLDSSGNGTADLSS
FKVSPPTSTLLFVGWSKTQNGAPLSESEIQALPVSSDISLYPVFKESYGVEFNTGDLSTGVTYIAPRRVLTGQPA
STIKPNDPTRPGYTFAGWYTAASGGAAFDFNQVLTKDTTLYAHWSPAQTTYTINYWQQSATDNKNATDAQKTYEY
AGQVTRSGLSLSNQTLTQQDINDKLPTGFKVNNTRTETSVMIKDDGSSVVNVYYDRKLITIKFAKYGGYSLPEYY
YSYNWSSDADTYTGLYGTTLAANGYQWKTGAWGYLANVGNNQVGTYGMSYLGEFILPNDTVDSDVIKLFP**K**GNIV
QTYRFF**K**QGLDGTYSLADTGGGAGADEFTFTEKYLGFNVKYYQRLYPDNYLFDQYASQTSAGVKVPISDEYYDRY
GAYHKDYLNLVVWYERNSYKIKYLDPLDNTELPNFPVKDVLYEQNLSSYAPDTTTVQP**K**PSRPGYVWDG**K**WYKDQ
AQTQVFDFNTTMPPHDVKVYAGWQKVTYRVNIDPNGGRLSKTDDTYLDLHYGDRIPDYTDITRDYIQDPSGTYYY
KYDSRDKDPDSTKDAYYTTDTSLSNVDTTTKYKYVKDAYKLVGWYYVNPDGSIRPYNFSGAVTQDINLRAIWRKA
GDYHIIYSNDAVGTDGKPALDASGQQLQTSNEPTDPDSYDDGSHSALLRRPTMPDGYRFRGWWYNGKIYNPYDSI
DIDAHLADANKNITIKPVIIPVGDIKLEDTSIKYNGNGGTRVENGNVVTQVETPRMELNSTTTIPENQYFTRTGY
NLIGWHHDKDLADTGRVEFTAGQSIGIDNNPDATNTLYAVWQPKEYTVRVSKTVVGLDEDKTKDFLFNPSETLQQ
ENFPLRDGQTKEFKVPYGTSISIDEQAYDEFKVSESITEKNLATGEADKTYDATGLQSLTVSGDVDISFTNTRIK
QKVRLQKVNVENDNNFLAGAVFDIYESDANGNKASHPMYSGLVTNDKGLLLVDANNYLSLPVGKYYLTETKAPPG
YLLPKNDISVLVISTGVTFEQNGNNATPIKENLVDGSTVYTFKITNSKGTELPSTGGIGTHIYILVGLALALPSG
LILYYRKKI

An E box containing a conserved glutamic residue has also been identified in 01524. The E box motif is underlined in SEQ ID NO: 36 below. The conserved glutamic acid (E), at amino acid residue 1344, is marked in bold. The E box motif, in particular the conserved glutamic acid residue, is thought to be important for the formation of oligomeric pilus-like structures of 01524. Preferred fragments of 01524 include the conserved glutamic acid residue. Preferably, fragments include the E box motif.

SEQ ID NO: 36
MLKKCQTFIIESLKKKKHPKEWKIIMWSLMILTTFLTTYFLILPAITVEETKTDDVGITLENKNSSQVTSSTSSS
QSSVEQSKPQTPASSVTETSSSEEAAYREEPLMFRGADYTVTVTLTKEAKIPKNADLKVTELKDNSATFKDYKKK
ALTEVAKQDSEIKNFKLYDITIESNGKEAEPQAPVKVEVNYDKPLEASDENLKVVHFKDDGQTEVLKSKDTAETK
NTSSDVAFKTDSFSIYAIVQEDNTEVPRLTYHFQNNDGTDYDFLTASGMQVHHQIIKDGESLGEVGIPTIKAGEH
FNGWYTYDPTTGKYGDPVKFGEPITVTETKEICVRPFMSKVATVTLYDDSAGKSILERYQVPLDSSGNGTADLSS
FKVSPPTSTLLFVGWSKTQNGAPLSESEIQALPVSSDISLYPVFKESYGVEFNTGDLSTGVTYIAPRRVLTGQPA
STIKPNDPTRPGYTFAGWYTAASGGAAFDFNQVLTKDTTLYAHWSPAQTTYTINYWQQSATDNKNATDAQKTYEY
AGQVTRSGLSLSNQTLTQQDINDKLPTGFKVNNTRTETSVMIKDDGSSVVNVYYDRKLITIKFAKYGGYSLPEYY
YSYNWSSDADTYTGLYGTTLAANGYQWKTGAWGYLANVGNNQVGTYGMSYLGEFILPNDTVDSDVIKLFPKGNIV
QTYRFFKQGLDGTYSLADTGGGAGADEFTFTEKYLGFNVKYYQRLYPDNYLFDQYASQTSAGVKVPISDEYYDRY
GAYHKDYLNLVVWYERNSYKIKYLDPLDNTELPNFPVKDVLYEQNLSSYAPDTTTVQPKPSRPGYVWDGKWYKDQ
AQTQVFDFNTTMPPHDVKVYAGWQKVTYRVNIDPNGGRLSKTDDTYLDLHYGDRIPDYTDITRDYIQDPSGTYYY
KYDSRDKDPDSTKDAYYTTDTSLSNVDTTTKYKYVKDAYKLVGWYYVNPDGSIRPYNFSGAVTQDINLRAIWRKA
GDYHIIYSNDAVGTDGKPALDASGQQLQTSNEPTDPDSYDDGSHSALLRRPTMPDGYRFRGWWYNGKIYNPYDSI
DIDAHLADANKNITIKPVIIPVGDIKLEDTSIKYNGNGGTRVENGNVVTQVETPRMELNSTTTIPENQYFTRTGY
NLIGWHHDKDLADTGRVEFTAGQSIGIDNNPDATNTLYAVWQPKEYTVRVSKTVVGLDEDKTKDFLFNPSETLQQ
ENFPLRDGQTKEFKVPYGTSISIDEQAYDEFKVSESITEKNLATGEADKTYDATGLQSLTVSGDVDISFTNTRIK
QKVRLQKVNVENDNNFLAGAVFDIYESDANGNKASHPMYSGLVTNDKGLLLVDANNYLSLPVGKYYLT**E**TKAPPG
YLLPKNDISVLVISTGVTFEQNGNNATPIKENLVDGSTVYTFKITNSKGTELPSTGGIGTHIYILVGLALALPSG
LILYYRKKI

01525

An example of an amino acid sequence for 01525 is set forth below. SEQ ID NO: 37 represents a 01525 sequence from GBS serotype III, strain isolate COH1.

SEQ ID NO: 37
MKRQISSDKLSQELDRVTYQKRFWSVIKNTIYILMAVASIAILIAVLWLPVLRIYGHSMNKTLSAGDVVFTVKGS
NFKTGDVVAFYYNNKVLVKRVIAESGDWVNIDSQGDVYVNQHKLKEPYVIHKALGNSNIKYPYQVPDKKIFVLGD
NRKTSIDSRSTSVGDVSEEQIVGKISFRIWPLGKISSIN

GBS 322

GBS 322 refers to a surface immunogenic protein, also referred to as "sip". Nucleotide and amino acid sequences of GBS 322 sequenced from serotype V isolated strain 2603 V/R are set forth in Ref. 3 as SEQ ID 8539 and SEQ ID 8540. These sequences are set forth below as SEQ ID NOS 38 and 39:

SEQ ID NO. 38
ATGAATAAAAAGGTACTATTGACATCGACAATGGCAGCTTCGCTATTATCAGTCGCAAGTGTTCAAGCACAAGAA
ACAGATACGACGTGGACAGCACGTACTGTTTCAGAGGTAAAGGCTGATTTGGTAAAGCAAGACAATAAATCATCA
TATACTGTGAAATATGGTGATACACTAAGCGTTATTTCAGAAGCAATGTCAATTGATATGAATGTCTTAGCAAAA
ATAAATAACATTGCAGATATCAATCTTATTTATCCTGAGACAACACTGACAGTAACTTACGATCAGAAGAGTCAT
ACTGCCACTTCAATGAAAATAGAAACACCAGCAACAAATGCTGCTGGTCAAACAACAGCTACTGTGGATTTGAAA
ACCAATCAAGTTTCTGTTGCAGACCAAAAAGTTTCTCTCAATACAATTTCGGAAGGTATGACACCAGAAGCAGCA
ACAACGATTGTTTCGCCAATGAAGACATATTCTTCTGCGCCAGCTTTGAAATCAAAAGAAGTATTAGCACAAGAG
CAAGCTGTTAGTCAAGCAGCAGCTAATGAACAGGTATCACCAGCTCCTGTGAAGTCGATTACTTCAGAAGTTCCA
GCAGCTAAAGAGGAAGTTAAACCAACTCAGACGTCAGTCAGTCAGTCAACAACAGTATCACCAGCTTCTGTTGCC
GCTGAAACACCAGCTCCAGTAGCTAAAGTAGCACCGGTAAGAACTGTAGCAGCCCCTAGAGTGGCAAGTGTTAAA
GTAGTCACTCCTAAAGTAGAAACTGGTGCATCACCAGAGCATGTATCAGCTCCAGCAGTTCCTGTGACTACGACT
TCACCAGCTACAGACAGTAAGTTACAAGCGACTGAAGTTAAGAGCGTTCCGGTAGCACAAAAAGCTCCAACAGCA
ACACCGGTAGCACAACCAGCTTCAACAACAAATGCAGTAGCTGCACATCCTGAAAATGCAGGGCTCCAACCTCAT
GTTGCAGCTTATAAAGAAAAAGTAGCGTCAACTTATGGAGTTAATGAATTCAGTACATACCGTGCGGGAGATCCA
GGTGATCATGGTAAAGGTTTAGCAGTTGACTTTATTGTAGGTACTAATCAAGCACTTGGTAATAAAGTTGCACAG
TACTCTACACAAAATATGGCAGCAAATAACATTTCATATGTTATCTGGCAACAAAAGTTTTACTCAAATACAAAC
AGTATTTATGGACCTGCTAATACTTGGAATGCAATGCCAGATCGTGGTGGCGTTACTGCCAACCACTATGACCAC

GTTCAGGTATCATTTAACAAATAATATAAAAAGGAAGCTATTTGGCTTCTTTTTTATATGCCTTGAATAGACTT
TCAAGGTTCTTATATAATTTTTATTA

SEQ ID NO. 39

MNKKVLLTSTMAASLLSVASVQAQETDTTWTARTVSEVKADLVKQDNKSSYTVKYGDTLSVISEAMSIDMNVLAK
INNIADINLIYPETTLTVTYDQKSHTATSMKIETPATNAAGQTTATVDLKTNQVSVADQKVSLNTISEGMTPEAA
TTIVSPMKTYSSAPALKSKEVLAQEQAVSQAAANEQVSPAPVKSITSEVPAAKEEVKPTQTSVSQSTTVSPASVA
AETPAPVAKVAPVRTVAAPRVASVKVVTPKVETGASPEHVSAPAVPVTTTSPATDSKLQATEVKSVPVAQKAPTA
TPVAQPASTTNAVAAHPENAGLQPHVAAYKEKVASTYGVNEFSTYRAGDPGDHGKGLAVDFIVGTNQALGNKVAQ
YSTQNMAANNISYVIWQQKFYSNTNSIYGPANTWNAMPDRGGVTANHYDHVHVSFNK

GBS 322 contains an N-terminal leader or signal sequence region which is indicated by the underlined sequence near the beginning of SEQ ID NO: 39. In one embodiment, one or more amino acids from the leader or signal sequence region of GBS 322 are removed. An example of such a GBS 322 fragment is set forth below as SEQ ID NO: 40.

SEQ ID NO: 40

DLVKQDNKSSYTVKYGDTLSVISEAMSIDMNVLAKINNIADINLIYPETTLTVTYDQKSHTATSMKIETPATNAA
GQTTATVDLKTNQVSVADQKVSLNTISEGMTPEAATTIVSPMKTYSSAPALKSKEVLAQEQAVSQAAANEQVSPA
PVKSITSEVPAAKEEVKPTQTSVSQSTTVSPASVAAETPAPVAKVAPVRTVAAPRVASVKVVTPKVETGASPEHV
SAPAVPVTTTSPATDSKLQATEVKSVPVAQKAPTATPVAQPASTTNAVAAHPENAGLQPHVAAYKEKVASTYGVN
EFSTYRAGDPGDHGKGLAVDFIVGTNQALGNKVAQYSTQNMAANNISYVIWQQKFYSNTNSIYGPANTWNAMPDR
GGVTANHYDHVHVSFNK

Additional preferred fragments of GBS 322 comprise the immunogenic epitopes identified in WO 03/068813, each of which are specifically incorporated by reference herein.

There may be an upper limit to the number of GBS proteins which will be in the compositions of the invention. Preferably, the number of GBS proteins in a composition of the invention is less than 20, less than 19, less than 18, less than 17, less than 16, less than 15, less than 14, less than 13, less than 12, less than 11, less than 10, less than 9, less than 8, less than 7, less than 6, less than 5, less than 4, or less than 3. Still more preferably, the number of GBS proteins in a composition of the invention is less than 6, less than 5, or less than 4. Still more preferably, the number of GBS proteins in a composition of the invention is 3.

The GBS proteins and polynucleotides used in the invention are preferably isolated, *i.e.*, separate and discrete, from the whole organism with which the molecule is found in nature or, when the polynucleotide or polypeptide is not found in nature, is sufficiently free of other biological macromolecules so that the polynucleotide or polypeptide can be used for its intended purpose.

Group A Streptococcus Adhesin Island Sequences

The GAS AI polypeptides of the invention can, of course, be prepared by various means (*e.g.* recombinant expression, purification from GAS, chemical synthesis *etc.*) and in various forms (*e.g.* native, fusions, glycosylated, non-glycosylated *etc.*). They are preferably prepared in substantially pure form (*i.e.* substantially free from other streptococcal or host cell proteins) or substantially isolated form.

The GAS AI proteins of the invention may include polypeptide sequences having sequence identity to the identified GAS proteins. The degree of sequence identity may vary depending on the amino acid sequence (a) in question, but is preferably greater than 50% (*e.g.* 60%, 65%, 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 99.5% or more). Polypeptides having sequence identity include homologs, orthologs, allelic variants and functional mutants of the identified GBS proteins. Typically, 50% identity or more between two proteins is considered to be an indication of functional equivalence. Identity between proteins is preferably determined by the Smith-Waterman homology search algorithm as implemented in the MPSRCH program (Oxford Molecular), using an affinity gap search with parameters *gap open penalty=12* and *gap extension penalty=1*.

The GAS adhesin island polynucleotide sequences may include polynucleotide sequences having sequence identity to the identified GAS adhesin island polynucleotide sequences. The degree of sequence identity may vary depending on the polynucleotide sequence in question, but is preferably greater than 50% (*e.g.* 60%, 65%, 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 99.5% or more).

The GAS adhesin island polynucleotide sequences of the invention may include polynucleotide fragments of the identified adhesin island sequences. The length of the fragment may vary depending on the polynucleotide sequence of the specific adhesin island sequence, but the fragment is preferably at least 10 consecutive polynucleotides, (*e.g.* at least 10, 12, 14, 16, 18, 20, 25, 30, 35, 40, 50, 60, 70, 80, 90, 100, 150, 200 or more).

The GAS adhesin island amino acid sequences of the invention may include polypeptide fragments of the identified GAS proteins. The length of the fragment may vary depending on the amino acid sequence of the specific GAS antigen, but the fragment is preferably at least 7 consecutive amino acids, (*e.g.* 8, 10, 12, 14, 16, 18, 20, 25, 30, 35, 40, 50, 60, 70, 80, 90, 100, 150, 200 or more). Preferably the fragment comprises one or more epitopes from the sequence. Other preferred fragments include (1) the N-terminal signal peptides of each identified GAS protein, (2) the identified GAS protein without their N-terminal signal peptides, and (3) each identified GAS protein wherein up to 10 amino acid residues (e.g. 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25 or more) are deleted from the N-terminus and/or the C-terminus *e.g.* the N-terminal amino acid residue may be deleted. Other fragments omit one or more domains of the protein (*e.g.* omission of a signal peptide, of a cytoplasmic domain, of a transmembrane domain, or of an extracellular domain).

GAS AI-1 sequences

As discussed above, a GAS AI-1 sequence is present in an M6 strain isolate (MGAS10394). Examples of GAS AI-1 sequences from M6 strain isolate MGAS10394 are set forth below.

M6_Spy0156: Spy0156 is a rofA transcriptional regulator. An example of an amino acid sequence for M6_Spy0156 is set forth in SEQ ID NO: 41.

SEQ ID NO: 41

```
MIEKYLESSIESKCQLVVLFFKTSYLPITEVAEKTGLTFLQLNHYCEELNAFFPDSLSMTIQKRMISCQFTHPFK
ETYLYQLYASSNVLQLLAFLIKNGSHSRPLTDFARSHFLSNSSAYRMREALIPLLRNFELKLSKNKIVGEEYRIR
YLIALLYSKFGIKVYDLTQQDKNTIHSFLSHSSTHLKTSPWLSESFSFYDILLALSWKRHQFSVTIPQTRIFQQL
KKLFIYDSLKKSSRDIIETYCQLNFSAGDLDYLYLIYITANNSFASLQWTPEHIRQCCQLFEENDTFRLLLKPII
TLLPNLKEQKPSLVKALMFFSKSFLFNLQHFIPETNLFVSPYYKGNQKLYTSLKLIVEEWLAKLPGKRYLNHKHF
HLFCHYVEQILRNIQPPLVVVFVASNFINAHLLTDSFPRYFSDKSIDFHSYIAR
```

M6_Spy0157: M6_Spy0157 is a fibronectin binding protein. It contains a sortase substrate motif LPXTG (SEQ ID NO: 122), shown in italics in the amino acid sequence SEQ ID NO: 42 .

SEQ ID NO: 42

MVSSYMFVRGEKMNNKIFLNKEASFLAHTKRKRRFAVTLVGVFFMLLACAGAIGFGQVAYAADEKTVPSHSSPNP
EFPWYGYDAYGKEYPGYNIWTRYHDLRVNLNGSRSYQVYCFNIQSNYPSQKNSFIKNWFKKIEGNGKSFVDYAHT
TKLGKEELEQRLLSLLYNAYPNDANGYMKGLEHLNAITVTQYAVWHYSDNSQYQFETLWESEAKEGKISRSQVTL
MREALKKLIDPNLEATAVNKIPSGYRLNIFESENEAYQNLLSAEYVPDDPPKPGETSEHNPKTPELDGTPIPEDP
KHPDDNLEPTLPPVMLDGEEVPEVPSESLEPALPPLMPELDGQEVPEKPSIDLPIEVPRYEFNNKDQSPLAGESG
ETEYITEVYGNQQNPVDIDKKLPNETGFSGNMVETEDTKEPEVLMGGQSESVEFTKDTQTGMSGQTTPQVETEDT
KEPEVLMGGQSESVEFTKDTQTGMSGQTTPQIETEDTKEPEVLMGGQSESVEFTKDTQTGMSGQTTPQIETEDTK
EPEVLMGGQSESVEFTKDTQTGMSGFSETATVVEDTRPKLVFHFDNNEPKVEENREKPTKNITPI*LPATG*DIENV
LAFLGILILSVLSIFSLLKNKQSNKKV

M6_Spy0157 contains an amino acid motif indicative of a cell wall anchor: **SEQ ID NO: 180** LPATG (shown in italics in SEQ ID NO: 42, above). In some recombinant host cell systems, it may be preferable to remove this motif to facilitate secretion of a recombinant M6_Spy0157 protein from the host cell. Alternatively, in other recombinant host cell systems, it may be preferable to use the cell wall anchor motif to anchor the recombinantly expressed protein to the cell wall. The extracellular domain of the expressed protein may be cleaved during purification or the recombinant protein may be left attached to either inactivated host cells or cell membranes in the final composition.

A pilin motif, discussed above, containing a conserved lysine (K) residue has also been identified in M6_Spy0157. The pilin motif sequence is underlined in SEQ ID NO: 42, below. Conserved lysine (K) residues are also marked in bold, at amino acid residues 277, 287, and 301. The pilin sequence, in particular the conserved lysine residues, are thought to be important for the formation of oligomeric, pilus-like structures. Preferred fragments of M6_Spy0157 include at least one conserved lysine residue. Preferably, fragments include the pilin sequence.

SEQ ID NO: 42

MVSSYMFVRGEKMNNKIFLNKEASFLAHTKRKRRFAVTLVGVFFMLLACAGAIGFGQVAYAADEKTVPSHSSPNP
EFPWYGYDAYGKEYPGYNIWTRYHDLRVNLNGSRSYQVYCFNIQSNYPSQKNSFIKNWFKKIEGNGKSFVDYAHT
TKLGKEELEQRLLSLLYNAYPNDANGYMKGLEHLNAITVTQYAVWHYSDNSQYQFETLWESEAKEGKISRSQVTL
MREALKKLIDPNLEATAVNKIPSGYRLNIFESENEAYQNLLSAEYVPDDPP**K**PGETSEHNP**K**TPELDGTPIPEDP
KHPDDNLEPTLPPVMLDGEEVPEVPSESLEPALPPLMPELDGQEVPEKPSIDLPIEVPRYEFNNKDQSPLAGESG
ETEYITEVYGNQQNPVDIDKKLPNETGFSGNMVETEDTKEPEVLMGGQSESVEFTKDTQTGMSGQTTPQVETEDT
KEPEVLMGGQSESVEFTKDTQTGMSGQTTPQIETEDTKEPEVLMGGQSESVEFTKDTQTGMSGQTTPQIETEDTK
EPEVLMGGQSESVEFTKDTQTGMSGFSETATVVEDTRPKLVFHFDNNEPKVEENREKPTKNITPILPATGDIENV
LAFLGILILSVLSIFSLLKNKQSNKKV

A repeated series of four E boxes containing a conserved glutamic residue have been identified in M6_Spy0157. The E-box motifs are underlined in SEQ ID NO: 42, below. The conserved glutamic acid (E) residues, at amino acid residues 415, 452, 489, and 526 are marked in bold. The E box motif, in particular the conserved glutamic acid residue, is thought to be important for the formation of oligomeric pilus-like structures of M6_Spy0157. Preferred fragments of M6_Spy0157 include at least one conserved glutamic acid residue. Preferably, fragments include at least one E box motif.

SEQ ID NO: 42

MVSSYMFVRGEKMNNKIFLNKEASFLAHTKRKRRFAVTLVGVFFMLLACAGAIGFGQVAYAADEKTVPSHSSPNP
EFPWYGYDAYGKEYPGYNIWTRYHDLRVNLNGSRSYQVYCFNIQSNYPSQKNSFIKNWFKKIEGNGKSFVDYAHT

TKLGKEELEQRLLSLLYNAYPNDANGYMKGLEHLNAITVTQYAVWHYSDNSQYQFETLWESEAKEGKISRSQVTL
MREALKKLIDPNLEATAVNKIPSGYRLNIFESENEAYQNLLSAEYVPDDPPKPGETSEHNPKTPELDGTPIPEDP
KHPDDNLEPTLPPVMLDGEEVPEVPSESLEPALPPLMPELDGQEVPEKPSIDLPIEVPRYEFNNKDQSPLAGESG
ETEYITEVYGNQQNPVDIDKKLPNETGFSGNMVETEDTK**E**PEVLMGGQSESVEFTKDTQTGMSGQTTPQVETEDT
K**E**PEVLMGGQSESVEFTKDTQTGMSGQTTPQIETEDTK**E**PEVLMGGQSESVEFTKDTQTGMSGQTTPQIETEDTK
EPEVLMGGQSESVEFTKDTQTGMSGFSETATVVEDTRPKLVFHFDNNEPKVEENREKPTKNITPILPATGDIENV
LAFLGILILSVLSIFSLLKNKQSNKKV

M6_Spy0158: M6_Spy0158 is a reverse transcriptase. An example of Spy0158 is shown in the amino acid sequence SEQ ID NO 43.

SEQ ID NO: 43
MSLRHQNKKGIRKEGWKSRPQSRWSDHCQLVAQKSVLKQAISKTVLAERGLFSCLDDYLERHALKVN

M6_Spy0159: M6_Spy0159 is a collagen adhesion protein. It contains a sortase substrate motif LPXSG, shown in italics in the amino acid sequence SEQ ID NO: 44.

SEQ ID NO: 44
MYSRLKRELVIVINRKKKYKLIRLMVTVGLIFSQLVLPIRRLGLQMISTQTKVIPQEIVTQTETQGTQVVATKQK
LESENSSLKVALKRESGFEHNATIDASLDTESQGDNSQRSVTQAIVTMALELRKQGLSIVDTKIVRIQSSTNQRN
DITTTLTFKNGLSLEGASTEANDPNVRVGIVNPNDTVQTITPTIKQDADGKVKNLVFTGRLGKQVIIVSTTRLKE
EQTISLDSYGELVIDGAVGLSQKDRPPYSKPITVNILKPKLSSIESSLDSKDFEIVKTIDNLYTWDDQFYLLDFI
SKQYEVLKTDYQSAKDSTPQTRDILFGEYTVEPLVMNKGHNNTINIYIRSTRPLGLKPIGAAPALIQPRSFRSLT
PRSTRMKRSAPVEKFEGELEHHKRIDYLGDNQNNPDTTIDDKEDEHDTSDLYRLYLDMTGKKNPLDILVVVDKSG
SMQEGIGSVQRYRYYAQRWDDYYSQWVYHGTFDYSSYQGESFNRGQIHYRYRGIVSVSDGIRRDDAVKNSLLGVN
GLLQRFVNINPENKLSVIGFQGSADYHAGKWYPDQSPRGGFYQPNLNNSRDAELLKGWSTNSLLDPNTLTALHNN
GTNYHAALLKAKEILNEVKDDGRRKIMIFISDGVPTFYFGEDGYRSGNGSSNDRNNVTRSQEGSKLAIDEFKARY
PNLSIYSLGVSKDINSDTASSSPVVLKYLSGEEHYYGITDTAELEKTLNKIVEDSKLSQLGISDSLSQYVDYYDK
QPDVLVTRKSKVNDETEILYQKDQVQEAGKDIIDKVVFTPKTTSQPKGKVTLTFKSDYKVDDEYTYTLSFNVKAS
DEAYEKYKDNEGRYSEMGDSDTDYGTNQTSSGKGGLPSNSDASVNYMADGREQKLPYKHPVIQVKTVPITFTKVD
ADNNQKKLAGVEFELRKEDKKIVWEKGTTGSNGQLNFKYLQKGKTYYLYETKAKLGYTLPENPWEVAVANNGDIK
VKHPIEGELKSKDGSYMIKNYKIYQ*LPSSG*GRGSQIFIIVGSMTATVALLFYRRQHRKKQY

M6_Spy0159 contains an amino acid motif indicative of a cell wall anchor: **SEQ ID NO: 181** LPSSG (shown in italics in SEQ ID NO: 44, above). In some recombinant host cell systems, it may be preferable to remove this motif to facilitate secretion of a recombinant M6_Spy0159 protein from the host cell. Alternatively, in other recombinant host cell systems, it may be preferable to use the cell wall anchor motif to anchor the recombinantly expressed protein to the cell wall. The extracellular domain of the expressed protein may be cleaved during purification or the recombinant protein may be left attached to either inactivated host cells or cell membranes in the final composition.

A pilin motif, discussed above, containing a conserved lysine (K) residue has also been identified in M6_Spy0159. The pilin motif sequence is underlined in SEQ ID NO: 44, below. Conserved lysine (K) residues are also marked in bold, at amino acid residues 265 and 276. The pilin sequence, in particular the conserved lysine residues, are thought to be important for the formation of oligomeric, pilus-like structures. Preferred fragments of M6_Spy0159 include at least one conserved lysine residue. Preferably, fragments include the pilin sequence.

SEQ ID NO: 44
MYSRLKRELVIVINRKKKYKLIRLMVTVGLIFSQLVLPIRRLGLQMISTQTKVIPQEIVTQTETQGTQVVATKQK
LESENSSLKVALKRESGFEHNATIDASLDTESQGDNSQRSVTQAIVTMALELRKQGLSIVDTKIVRIQSSTNQRN
DITTTLTFKNGLSLEGASTEANDPNVRVGIVNPNDTVQTITPTIKQDADGKVKNLVFTGRLGKQVIIVSTTRLKE
EQTISLDSYGELVIDGAVGLSQKDRPPYSKPITVNILKP**K**LSSIESSLDS**K**DFEIVKTIDNLYTWDDQFYLLDFI
SKQYEVLKTDYQSAKDSTPQTRDILFGEYTVEPLVMNKGHNNTINIYIRSTRPLGLKPIGAAPALIQPRSFRSLT
PRSTRMKRSAPVEKFEGELEHHKRIDYLGDNQNNPDTTIDDKEDEHDTSDLYRLYLDMTGKKNPLDILVVVDKSG

SMQEGIGSVQRYRYYAQRWDDYYSQWVYHGTFDYSSYQGESFNRGQIHYRYRGIVSVSDGIRRDDAVKNSLLGVN
GLLQRFVNINPENKLSVIGFQGSADYHAGKWYPDQSPRGGFYQPNLNNSRDAELLKGWSTNSLLDPNTLTALHNN
GTNYHAALLKAKEILNEVKDDGRRKIMIFISDGVPTFYFGEDGYRSGNGSSNDRNNVTRSQEGSKLAIDEFKARY
PNLSIYSLGVSKDINSDTASSSPVVLKYLSGEEHYYGITDTAELEKTLNKIVEDSKLSQLGISDSLSQYVDYYDK
QPDVLVTRKSKVNDETEILYQKDQVQEAGKDIIDKVVFTPKTTSQPKGKVTLTFKSDYKVDDEYTYTLSFNVKAS
DEAYEKYKDNEGRYSEMGDSDTDYGTNQTSSGKGGLPSNSDASVNYMADGREQKLPYKHPVIQVKTVPITFTKVD
ADNNQKKLAGVEFELRKEDKKIVWEKGTTGSNGQLNFKYLQKGKTYYLYETKAKLGYTLPENPWEVAVANNGDIK
VKHPIEGELKSKDGSYMIKNYKIYQLPSSGGRGSQIFIIVGSMTATVALLFYRRQHRKKQY

An E box containing a conserved glutamic residue has been identified in M6_Spy0159. The E-box motif is underlined in SEQ ID NO: 44, below. The conserved glutamic acid (E), at amino acid residue 950, is marked in bold. The E box motif, in particular the conserved glutamic acid residue, is thought to be important for the formation of oligomeric pilus-like structures of M6_Spy0159. Preferred fragments of M6_Spy0159 include the conserved glutamic acid residue. Preferably, fragments include the E box motif.

SEQ ID NO: 44
MYSRLKRELVIVINRKKKYKLIRLMVTVGLIFSQLVLPIRRLGLQMISTQTKVIPQEIVTQTETQGTQVVATKQK
LESENSSLKVALKRESGFEHNATIDASLDTESQGDNSQRSVTQAIVTMALELRKQGLSIVDTKIVRIQSSTNQRN
DITTTLTFKNGLSLEGASTEANDPNVRVGIVNPNDTVQTITPTIKQDADGKVKNLVFTGRLGKQVIIVSTTRLKE
EQTISLDSYGELVIDGAVGLSQKDRPPYSKPITVNILKPKLSSIESSLDSKDFEIVKTIDNLYTWDDQFYLLDFI
SKQYEVLKTDYQSAKDSTPQTRDILFGEYTVEPLVMNKGHNNTINIYIRSTRPLGLKPIGAAPALIQPRSFRSLT
PRSTRMKRSAPVEKFEGELEHHKRIDYLGDNQNNPDTTIDDKEDEHDTSDLYRLYLDMTGKKNPLDILVVVDKSG
SMQEGIGSVQRYRYYAQRWDDYYSQWVYHGTFDYSSYQGESFNRGQIHYRYRGIVSVSDGIRRDDAVKNSLLGVN
GLLQRFVNINPENKLSVIGFQGSADYHAGKWYPDQSPRGGFYQPNLNNSRDAELLKGWSTNSLLDPNTLTALHNN
GTNYHAALLKAKEILNEVKDDGRRKIMIFISDGVPTFYFGEDGYRSGNGSSNDRNNVTRSQEGSKLAIDEFKARY
PNLSIYSLGVSKDINSDTASSSPVVLKYLSGEEHYYGITDTAELEKTLNKIVEDSKLSQLGISDSLSQYVDYYDK
QPDVLVTRKSKVNDETEILYQKDQVQEAGKDIIDKVVFTPKTTSQPKGKVTLTFKSDYKVDDEYTYTLSFNVKAS
DEAYEKYKDNEGRYSEMGDSDTDYGTNQTSSGKGGLPSNSDASVNYMADGREQKLPYKHPVIQVKTVPITFTKVD
ADNNQKKLAGVEFELRKEDKKIVWEKGTTGSNGQLNFKYLQKGKTYYLY**E**TKAKLGYTLPENPWEVAVANNGDIK
VKHPIEGELKSKDGSYMIKNYKIYQLPSSGGRGSQIFIIVGSMTATVALLFYRRQHRKKQY

M6_Spy0160: M6_Spy0160 is a fimbrial structural subunit. It contains a sortase substrate motif LPXTG (SEQ ID NO: 122), shown in italics in amino acid sequence SEQ ID NO: 45.

SEQ ID NO: 45
MTNRRETVREKILITAKKLMLACLAILAVVGLGMTRVSALSKDDTAQLKITNIEGGPTVTLYKIGEGVYNTNGDS
FINFKYAEGVSLTETGPTSQEITTIANGINTGKIKPFSTENVSISNGTATYNARGASVYIALLTGATDGRTYNPI
LLAASYNGEGNLVTKNIDSKSNYLYGQTSVAKSSLPSITKKVTGTIDDVNKKTTSLGSVLSYSLTFELPSYTKEA
VNKTVYVSDNMSEGLTFNFNSLTVEWKGKMANITEDGSVMVENTKIGIAKEVNNGFNLSFIYDSLESISPNISYK
AVVNNKAIVGEEGNPNKAEFFYSNNPTKGNTYDNLDKKPDKGNGITSKEDSKIVYTYQIAFRKVDSVSKTPLIGA
IFGVYDTSNKLIDIVTTNKNGYAISTQVSSGKYKIKELKAPKGYSLNTETYEITANWVTATVKTSANSKSTTYTS
DKNKATDNSEQVGWLKNGIFYSIDSRPTGNDVKEAYIESTKALTDGTTFSKSNEGSGTVLLETDIPNTKLGE*LPS*
*TG*SIGTYLFKAIGSAAMIGAIGIYIVKRRKA

M6_Spy0160 contains an amino acid motif indicative of a cell wall anchor: **SEQ ID NO: 131** LPSTG (shown in italics in SEQ ID NO: 45, above). In some recombinant host cell systems, it may be preferable to remove this motif to facilitate secretion of a recombinant M6_Spy0160 protein from the host cell. Alternatively, in other recombinant host cell systems, it may be preferable to use the cell wall anchor motif to anchor the recombinantly expressed protein to the cell wall. The extracellular domain of the expressed protein may be cleaved during purification or the recombinant protein may be left attached to either inactivated host cells or cell membranes in the final composition.

An E box containing a conserved glutamic residue has been identified in M6_Spy0160. The E-box motif is underlined in SEQ ID NO: 45, below. The conserved glutamic acid (E), at amino acid residue 412, is marked in bold. The E box motif, in particular the conserved glutamic acid residue, is thought to be important for the formation of oligomeric pilus-like structures of M6_Spy0160. Preferred fragments of M6_Spy0160 include the conserved glutamic acid residue. Preferably, fragments include the E box motif.

SEQ ID NO: 45
MTNRRETVREKILITAKKLMLACLAILAVVGLGMTRVSALSKDDTAQLKITNIEGGPTVTLYKIGEGVYNTNGDS
FINFKYAEGVSLTETGPTSQEITTIANGINTGKIKPFSTENVSISNGTATYNARGASVYIALLTGATDGRTYNPI
LLAASYNGEGNLVTKNIDSKSNYLYGQTSVAKSSLPSITKKVTGTIDDVNKKTTSLGSVLSYSLTFELPSYTKEA
VNKTVYVSDNMSEGLTFNFNSLTVEWKGKMANITEDGSVMVENTKIGIAKEVNNGFNLSFIYDSLESISPNISYK
AVVNNKAIVGEEGNPNKAEFFYSNNPTKGNTYDNLDKKPDKGNGITSKEDSKIVYTYQIAFRKVDSVSKTPLIGA
IFGVYDTSNKLIDIVTTNKNGYAISTQVSSGKYKIK**E**LKAPKGYSLNTETYEITANWVTATVKTSANSKSTTYTS
DKNKATDNSEQVGWLKNGIFYSIDSRPTGNDVKEAYIESTKALTDGTTFSKSNEGSGTVLLETDIPNTKLGELPS
TGSIGTYLFKAIGSAAMIGAIGIYIVKRRKA

M6_Spy0161 is a srtB type sortase. An example of an amino acid sequence of M6_Spy-161 is shown in SEQ ID NO: 46.

SEQ ID NO: 46
MTERLKNLGILLLFLLGTAIFLYPTLSSQWNAYRDRQLLSTYHKQVIQKKPSEMEEVWQKAKAYNARLGIQPVPD
AFSFRDGIHDKNYESLLQIENNDIMGYVEVPSIKVTLPIYHYTTDEVLTKGAGHLFGSALPVGGDGTHTVISAHR
GLPSAEMFTNLNLVKKGDTFYFRVLNKVLAYKVDQILIVEPDQATSLSGVMGKDYATLVTCTPYGVNTKRLLVRG
HRIAYHYKKYQQAKKAMKLVDKSRMWAEVVCAAFGVVIAIILVFMYSRVSAKKSK

As discussed above, applicants have also determined the nucleotide and encoded amino acid sequence of fimbrial structural subunits in several other GAS AI-1 strains of bacteria. Examples of sequences of these fimbrial structural subunits are set forth below.

M6 strain isolate CDC SS 410 is a GAS AI-1 strain of bacteria. CDC SS 410_fimbrial is thought to be a fimbrial structural subunit of M6 strain isolate CDC SS 410. An example of a nucleotide sequence encoding the CDC SS 410_fimbrial protein (SEQ ID NO: 267) and a CDC SS 410_fimbrial protein amino acid sequence (SEQ ID NO: 268) are set forth below.

SEQ ID NO: 267

aaagatgatactgcacaactaaagataacaaatattgaaggtgggccaacagtaacactt
tataaaataggagaaggtgtttacaacactaatggtgattcttttattaactttaaatat
gctgaggggtttctttaactgaaacaggacctacatcacaagaaattactactattgca
aatggtattaatacgggtaaaataaagccttttagtactgaaaacgttagtatttctaat
ggaacagcaacttataatgcgagaggtgcatctgtttatattgcattattaacaggtgcg
acagatggccgtacctacaatcctattttattagctgcatcttataatggtgagggaaat
ttagttactaaaatattgattccaaatctaattatttatatggacaaacaagtgttgca
aaatcatcattaccatctattacaaagaaagtaaccgggacaatagatgacgtgaataaa
aagactacctcgttaggaagtgtattgtcttattcgctgacatttgaattaccaagttat
accaaagaagcagtcaataaaacagtatatgtttctgataatatgtcggaaggtcttact
tttaactttaatagtcttacagtagaatggaaaggtaagatggctaatattactgaagat
ggttcagtaatggtagaaaatacaaaaatcggaatagctaaggaggttaataacggtttt
aatttaagttttatttatgatagtttagaatctatatcaccaatataagttataaagct
gttgtaaacaataaagctattgttggtgaagagggtaatcctaataaagctgaattcttc
tattcaaataatccaacaaaaggtaatacatacgataatttagataagaagcctgataaa
gggaatggtattacatccaaagaagattctaaaattgtttatacttatcaaatagcgttt
agaaaagttgatagtgttagtaagaccccacttattggtgcaatttttggagtttatgat
actagtaataaattaattgatattgttacaaccaataaaaatggatatgctatttcaaca

caagtatcttcaggaaaatataaaattaaggaattaaaagctcctaaaggttattcattg
aatacagaaacttatgaaattacggcaaattgggtaactgctacagtcaagacaagtgct
aattcaaaaagtactacttatacatctgataaaaataaggcgacagataattcagagcaa
gtaggatggttaaaaaatggtatattctattctatagatagtagacctacaggaaatgat
gttaaagaggcttatattgaatctactaaggctttaactgatggaacaactttctcaaaa
tcgaatgaaggttcaggtacagtattattagaaactgacatccctaacaccaagctaggt
gaactc

SEQ ID NO: 268

KDDTAQLKITNIEGGPTVTLYKIGEGVYNTNGDSFINFKYAEGV
SLTETGPTSQEITTIANGINTGKIKPFSTENVSISNGTATYNARGASVYIALLTGATD
GRTYNPILLAASYNGEGNLVTKNIDSKSNYLYGQTSVAKSSLPSITKKVTGTIDDVNK
KTTSLGSVLSYSLTFELPSYTKEAVNKTVYVSDNMSEGLTFNFNSLTVEWKGKMANIT
EDGSVMVENTKIGIAKEVNNGFNLSFIYDSLESISPNISYKAVVNNKAIVGEEGNPNK
AEFFYSNNPTKGNTYDNLDKKPDKGNGITSKEDSKIVYTYQIAFRKVDSVSKTPLIGA
IFGVYDTSNKLIDIVTTNKNGYAISTQVSSGKYKIKELKAPKGYSLNTETYEITANWV
TATVKTSANSKSTTYTSDKNKATDNSEQVGWLKNGIFYSIDSRPTGNDVKEAYIESTK
ALTDGTTFSKSNEGSGTVLLETDIPNTKLGEL

M6 strain isolate ISS 3650 is a GAS AI-1 strain of bacteria. ISS3650_fimbrial is thought to be a fimbrial structural subunit of M6 strain isolate ISS 3650. An example of a nucleotide sequence encoding the ISS3650_fimbrial protein (SEQ ID NO: 269) and an ISS3650_fimbrial protein amino acid sequence (SEQ ID NO: 270) are set forth below.

SEQ ID NO: 269

gaatggaaaggtaagatggctaatattactgaagatggttcagtaatggtagaaaataca
aaaatcggaatagctaaggaggttaataacggttttaatttaagttttatttatgatagt
ttagaatctatatcaccaaatataagttataaagctgttgtaaacaataaagctattgtt
ggtgaagagggtaatcctaataaagctgaattcttctattcaaataatccaacaaaaggt
aatacatacgataatttagataagaagcctgataaagggaatggtattacatccaaagaa
gattctaaaattgtttatacttatcaaatagcgtttagaaaagttgatagtgttagtaag
accccacttattggtgcaattttggagtttatgatactagtaataaattaattgatatt
gttacaaccaataaaaatggatatgctatttcaacacaagtatcttcaggaaaatataaa
attaaggaattaaaagctcctaaaggttattcattgaatacagaaacttatgaaattacg
gcaaattgggtaactgctacagtcaagacaagtgctaattcaaaaagtactacttataca
tctgataaaaataaggcgacagataattcagagcaagtaggatggttaaaaaatggtata
ttctattctatagatagtagacctacaggaaatgatgttaaagaggcttatattgaatct
actaaggctttaactgatggaacaactttctcaaaatcgaatgaaggttcaggtacagta
ttattagaaactgacatcc

SEQ ID NO: 270

EWKGKMANITEDGSVMVENTKIGIAKEVNNGFNLSFIYDSLESI
SPNISYKAVVNNKAIVGEEGNPNKAEFFYSNNPTKGNTYDNLDKKPDKGNGITSKEDS
KIVYTYQIAFRKVDSVSKTPLIGAIFGVYDTSNKLIDIVTTNKNGYAISTQVSSGKYK
IKELKAPKGYSLNTETYEITANWVTATVKTSANSKSTTYTSDKNKATDNSEQVGWLKN
GIFYSIDSRPTGNDVKEAYIESTKALTDGTTFSKSNEGSGTVLLETDI

M23 strain isolate DSM2071 is a GAS AI-1 strain of bacteria. DSM2071_fimbrial is thought to be a fimbrial structural subunit of M23 strain DSM2071. An example of a nucleotide sequence encoding the DSM2071_fimbrial protein (SEQ ID NO: 251) and a DSM2071_fimbrial protein amino acid sequence (SEQ ID NO: 252) are set forth below.

SEQ ID NO: 251

atgagagagaaaatattaatagcagcaaaaaaactaatgctagcttgtttagctatctta
gctgtagtagggcttggaatgacaagagtatcagctttatcaaaagatgataaggcggag
ttgaagataacaaatatcgaaggtaaaccgaccgtgacactgtataaaattggtgatgga
aaatacagtgagcgaggggattcttttattggatttgagttaaagcaaggtgtggagcta
aataaggcaaaacctacatctcaagaaataaataaaatcgctaatggtattaataaaggt
agtgttaaggctgaagtagttaatataaaagaacatgctagtacaacttatagttataca

acaactggtgcaggtatttacttggctatattgactggagctactgatggacgtgcctat
aatcctatcttactgacagcttcttacaatgaggaaaatccacttaagggagggcagatt
gacgcaactagtcattatctttttggagaagaagcagttgctaaatctagccaaccaaca
attagcaagtcaattacaaaatccacaaaagatggtgataaagatacagcatctgtaggt
gaaaaagttgattacaaattaactgttcagttaccaagttattcgaaagatgctatcaat
aaaacggtgtttatcactgacaaattgtctcagggacttactttccttccaaaaagttta
aagattatctggaatggtcaaacgttaacaaaggtgaatgaagaatttaaagctggagat
aaggtaattgctcaacttaaggttgaaaataatggatttaatctgaactttaattatgat
aaccttgataatcatgccccagaagttaactatagtgctctactaaatgaaaacgcagtt
gttggtaaaggtggtaatgacaataatgtagactattactattcaaataatccgaataaa
ggagagacccataaaacaactgagaagcctaaagagggtgaaggtactggtatcactaaa
aagacggataaaaaaaccgtctacacctatcgtgtagcctttaagaaaacaggcaaagat
catgccccactagctggtgctgttttcggtatctattcagataaggaagcgaaacaatta
gtcgatattgttgtgacaaatgcacagggttatgcagcatcaagcgaagttgggaaaggg
acttattacattaaagaaattaaatcccctaagggttactctttaaatacaaatatttat
gaagtggaaacttcatgggaaaaagctacaacgacttctacaactaatcgtttagagaca
atttatacaacagatgataatcaaaagtctccaggaactaatacagttggttggttggaa
gatggtgtcttttacaaagaaaatccaggtggtgatgctaaacttgcctatatcaaacaa
tcaacagaggagacttctacaactatagaagtcaaagaaaatcaagctgaaggttcaggt
acggtattattagaaactgaaattcctaacaccaaattaggtgaattaccttcgacaggt
agcattggtacttacctctttaaagctattggttcggctgctatgatcggtgcaattggt
atttatattgttaaacgtcgtaaagcttaa

SEQ ID NO: 252

MREKILIAAKKLMLACLAILAVVGLGMTRVSALSKDDKAELKIT
NIEGKPTVTLYKIGDGKYSERGDSFIGFELKQGVELNKAKPTSQEINKIANGINKGSV
KAEVVNIKEHASTTYSYTTTGAGIYLAILTGATDGRAYNPILLTASYNEENPLKGGQI
DATSHYLFGEEAVAKSSQPTISKSITKSTKDGDKDTASVGEKVDYKLTVQLPSYSKDA
INKTVFITDKLSQGLTFLPKSLKIIWNGQTLTKVNEEFKAGDKVIAQLKVENNGFNLN
FNYDNLDNHAPEVNYSALLNENAVVGKGGNDNNVDYYYSNNPNKGETHKTTEKPKEGE
GTGITKKTDKKTVYTYRVAFKKTGKDHAPLAGAVFGIYSDKEAKQLVDIVVTNAQGYA
ASSEVGKGTYYIKEIKSPKGYSLNTNIYEVETSWEKATTTSTTNRLETIYTTDDNQKS
PGTNTVGWLEDGVFYKENPGGDAKLAYIKQSTEETSTTIEVKENQAEGSGTVLLETEI
PNTKLGELPSTGSIGTYLFKAIGSAAMIGAIGIYIVKRRKA

GAS AI-2 sequences

As discussed above, a GAS AI-2 sequence is present in an M1 strain isolate (SF370). Examples of GAS AI-2 sequences from M1 strain isolate SF370 are set forth below.

Spy0124 is a rofA transcriptional regulator. An example of an amino acid sequence for Spy0124 is set forth in SEQ ID NO:47.

SEQ ID NO: 47

MIEKYLESSIESKCQLIVLFFKTSYLPITEVAEKTGLTFLQLNHYCEELNAFFPGSLSMTIQKRMISCQFTHPFK
ETYLYQLYASSNVLQLLAFLIKNGSHSRPLTDFARSHFLSNSSAYRMREALIPLLRNFELKLSKNKIVGEEYRIR
YLIALLYSKFGIKVYDLTQQDKNTIHSFLSHSSTHLKTSPWLSESFSFYDILLALSWKRHQFSVTIPQTRIFQQL
KKLFVYDSLKKSSHDIIETYCQLNFSAGDLDYLYLIYITANNSFASLQWTPEHIRQYCQLFEENDTFRLLLNPII
TLLPNLKEQKASLVKALMFFSKSFLFNLQHFIPETNLFVSPYYKGNQKLYTSLKLIVEEWMAKLPGKRDLNHKHF
HLFCHYVEQSLRNIQPPLVVVFVASNFINAHLLTDSFPRYFSDKSIDFHSYYLLQDNVYQIPDLKPDLVITHSQL
IPFVHHELTKGIAVAEISFDESILSIQELMYQVKEEKFQADLTKQLT

GAS 015 is also referred to as Cpa. It contains a sortase substrate motif VVXTG (SEQ ID NO: 135), shown in italics in SEQ ID NO: 48.

SEQ ID NO: 48

LRGEKMKKTRFPNKLNTLNTQRVLSKNSKRFTVTLVGVFLMIFALVTSMVGAKTVFGLVESSTPNAINPDSSSEY
RWYGYESYVRGHPYYKQFRVAHDLRVNLEGSRSYQVYCFNLKKAFPLGSDSSVKKWYKKHDGISTKFEDYAMSPR
ITGDELNQKLRAVMYNGHPQNANGIMEGLEPLNAIRVTQEAVWYYSDNAPISNPDESFKRESESNLVSTSQLSLM
RQALKQLIDPNLATKMPKQVPDDFQLSIFESEDKGDKYNKGYQNLLSGGLVPTKPPTPGDPPMPPNQPQTTSVLI
RKYAIGDYSKLLEGATLQLTGDNVNSFQARVFSSNDIGERIELSDGTYTLTELNSPAGYSIAEPITFKVEAGKVY

TIIDGKQIENPNKEIVEPYSVEAYNDFEEFSVLTTQNYAKFYYAKNKNGSSQVVYCFNADLKSPPDSEDGGKTMT
PDFTTGEVKYTHIAGRDLFKYTVKPRDTDPDTFLKHIKKVIEKGYREKGQAIEYSGLTETQLRAATQLAIYYFTD
SAELDKDKLKDYHGFGDMNDSTLAVAKILVEYAQDSNPPQLTDLDFFIPNNNKYQSLIGTQWHPEDLVDIIRMED
KKEVIPVTHNLTLRKTVTGLAGDRTKDFHFEIELKNNKQELLSQTVKTDKTNLEFKDGKATINLKHGESLTLQGL
PEGYSYLVKETDSEGYKVKVNSQEVANATVSKTGITSDETLAFENNKEP*VVPTG*VDQKINGYLALIVIAGISLGI
WGIHTIRIRKHD

GAS 015 contains an amino acid motif indicative of a cell wall anchor: **SEQ ID NO: 182** VVPTG (shown in italics in SEQ ID NO: 48, above). In some recombinant host cell systems, it may be preferable to remove this motif to facilitate secretion of a recombinant GAS 015 protein from the host cell. Alternatively, in other recombinant host cell systems, it may be preferable to use the cell wall anchor motif to anchor the recombinantly expressed protein to the cell wall. The extracellular domain of the expressed protein may be cleaved during purification or the recombinant protein may be left attached to either inactivated host cells or cell membranes in the final composition.

A pilin motif, discussed above, containing a conserved lysine (K) residue has also been identified in GAS 015. The pilin motif sequence is underlined in SEQ ID NO: 48, below. Conserved lysine (K) residues are also marked in bold, at amino acid residue 243. The pilin sequence, in particular the conserved lysine residues, are thought to be important for the formation of oligomeric, pilus-like structures. Preferred fragments of GAS 015 include the conserved lysine residue. Preferably, fragments include the pilin sequence.

SEQ ID NO: 48
LRGEKMKKTRFPNKLNTLNTQRVLSKNSKRFTVTLVGVFLMIFALVTSMVGAKTVFGLVESSTPNAINPDSSSEY
RWYGYESYVRGHPYYKQFRVAHDLRVNLEGSRSYQVYCFNLKKAFPLGSDSSVKKWYKKHDGISTKFEDYAMSPR
ITGDELNQKLRAVMYNGHPQNANGIMEGLEPLNAIRVTQEAVWYYSDNAPISNPDESFKRESESNLVSTSQLSLM
RQALKQLIDPNLATKMPKQVPDDFQLSIFESEDKGDKYNKGYQNLLSGGLVPTKPPTPGDPPMPPNQPQTTSVLI
RKYAIGDYSKLLEGATLQLTGDNVNSFQARVFSSNDIGERIELSDGTYTLTELNSPAGYSIAEPITFKVEAGKVY
TIIDGKQIENPNKEIVEPYSVEAYNDFEEFSVLTTQNYAKFYYAKNKNGSSQVVYCFNADLKSPPDSEDGGKTMT
PDFTTGEVKYTHIAGRDLFKYTVKPRDTDPDTFLKHIKKVIEKGYREKGQAIEYSGLTETQLRAATQLAIYYFTD
SAELDKDKLKDYHGFGDMNDSTLAVAKILVEYAQDSNPPQLTDLDFFIPNNNKYQSLIGTQWHPEDLVDIIRMED
KKEVIPVTHNLTLRKTVTGLAGDRTKDFHFEIELKNNKQELLSQTVKTDKTNLEFKDGKATINLKHGESLTLQGL
PEGYSYLVKETDSEGYKVKVNSQEVANATVSKTGITSDETLAFENNKEPVVPTGVDQKINGYLALIVIAGISLGI
WGIHTIRIRKHD

An E box containing a conserved glutamic residue has been identified in GAS 015. The E-box motif is underlined in SEQ ID NO: 48, below. The conserved glutamic acid (E), at amino acid residue 352, is marked in bold. The E box motif, in particular the conserved glutamic acid residue, is thought to be important for the formation of oligomeric pilus-like structures of GAS 015. Preferred fragments of GAS 015 include the conserved glutamic acid residue. Preferably, fragments include the E box motif.

SEQ ID NO: 48
LRGEKMKKTRFPNKLNTLNTQRVLSKNSKRFTVTLVGVFLMIFALVTSMVGAKTVFGLVESSTPNAINPDSSSEY
RWYGYESYVRGHPYYKQFRVAHDLRVNLEGSRSYQVYCFNLKKAFPLGSDSSVKKWYKKHDGISTKFEDYAMSPR
ITGDELNQKLRAVMYNGHPQNANGIMEGLEPLNAIRVTQEAVWYYSDNAPISNPDESFKRESESNLVSTSQLSLM
RQALKQLIDPNLATKMPKQVPDDFQLSIFESEDKGDKYNKGYQNLLSGGLVPTKPPTPGDPPMPPNQPQTTSVLI
RKYAIGDYSKLLEGATLQLTGDNVNSFQARVFSSNDIGERIELSDGTYTLTELNSPAGYSIAEPITFKVEAGKVY
TIIDGKQIENPNKEIVEPYSVEAYNDFEEFSVLTTQNYAKFYYAKNKNGSSQVVYCFNADLKSPPDSEDGGKTMT
PDFTTGEVKYTHIAGRDLFKYTVKPRDTDPDTFLKHIKKVIEKGYREKGQAIEYSGLTETQLRAATQLAIYYFTD
SAELDKDKLKDYHGFGDMNDSTLAVAKILVEYAQDSNPPQLTDLDFFIPNNNKYQSLIGTQWHPEDLVDIIRMED
KKEVIPVTHNLTLRKTVTGLAGDRTKDFHFEIELKNNKQELLSQTVKTDKTNLEFKDGKATINLKHGESLTLQGL

PEGYSYIVKETDSEGYKVKVNGQEVANATVSKTGITSDETLAFENNKEPVVPTGVDQKINGYLALIVIAGISLGI
WGIHTIRIRKHD

Spy0127 is a LepA putative signal peptidase. An example of an amino acid sequence for Spy0127 is set forth in SEQ ID NO: 49.

SEQ ID NO: 49

MIIKRNDMAPSVKAGDAILFYRLSQTYKVEEAVVYEDSKTSITKVGRIIAQAGDEVDLTEQGELKINGHIQNEGL
TFIKSREANYPYRIADNSYLILNDYYSQESENYLQDAIAKDAIKGTINTLIRLRNH

Spy0128 is thought to be a fibrial protein. It contains a sortase substrate motif EVXTG (SEQ ID NO: 136) shown in italics in SEQ ID NO: 50.

SEQ ID NO: 50

MKLRHLLLTGAALTSFAATTVHGETVVNGAKLTVTKNLDLVNSNALIPNTDFTFKIEPDTTVNEDGNKFKGVALN
TPMTKVTYTNSDKGGSNTKTAEFDFSEVTFEKPGVYYYKVTEEKIDKVPGVSYDTTSYTVQVHVLWNEEQQKPVA
TYIVGYKEGSKVPIQFKNSLDSTTLTVKKKVSGTGGDRSKDFNFGLTLKANQYYKASEKVMIEKTTKGGQAPVQT
EASIDQLYHFTLKDGESIKVTNLPVGVDYVVTEDDYKSEKYTTNVEVSPQDGAVKNIAGNSTEQETSTDKDMTIT
FTNKKDF*EVPTG*VAMTVAPYIALGIVAVGGALYFVKKKNA

Spy0128 contains an amino acid motif indicative of a cell wall anchor: **SEQ ID NO: 183** EVPTG (shown in italics in SEQ ID NO: 50, above). In some recombinant host cell systems, it may be preferable to remove this motif to facilitate secretion of a recombinant Spy0128 protein from the host cell. Alternatively, in other recombinant host cell systems, it may be preferable to use the cell wall anchor motif to anchor the recombinantly expressed protein to the cell wall. The extracellular domain of the expressed protein may be cleaved during purification or the recombinant protein may be left attached to either inactivated host cells or cell membranes in the final composition.

Two E boxes containing a conserved glutamic residue have been identified in Spy0128. The E-box motifs are underlined in SEQ ID NO: 50, below. The conserved glutamic acid (E) residues, at amino acid residues 271 and 290, are marked in bold. The E box motifs, in particular the conserved glutamic acid residues, are thought to be important for the formation of oligomeric pilus-like structures of Spy0128. Preferred fragments of Spy0128 include at least one conserved glutamic acid residue. Preferably, fragments include at least one E box motif.

SEQ ID NO: 50

MKLRHLLLTGAALTSFAATTVHGETVVNGAKLTVTKNLDLVNSNALIPNTDFTFKIEPDTTVNEDGNKFKGVALN
TPMTKVTYTNSDKGGSNTKTAEFDFSEVTFEKPGVYYYKVTEEKIDKVPGVSYDTTSYTVQVHVLWNEEQQKPVA
TYIVGYKEGSKVPIQFKNSLDSTTLTVKKKVSGTGGDRSKDFNFGLTLKANQYYKASEKVMIEKTTKGGQAPVQT
EASIDQLYHFTLKDGESIKVTNLPVGVDYVVTEDDYKSEKYTTNV**E**VSPQDGAVKNIAGNSTEQ**E**TSTDKDMTIT
FTNKKDFEVPTGVAMTVAPYIALGIVAVGGALYFVKKKNA

Spy0129 is a srtC1 type sortase. An example of an amino acid sequence for Spy0129 is set forth in SEQ ID NO: 51.

SEQ ID NO: 51

MIVRLIKLLDKLINVIVLCFFFLCLLIAALGIYDALTVYQGANATNYQQYKKKGVQFDDLLAINSDVMAWLTVKG
THIDYPIVQGENNLEYINKSVEGEYSLSGSVFLDYRNKVTFEDKYSLIYAHHMAGNVMFGELPNFRKKSFFNKHK
EFSIETKTKQKLKINIFACIQTDAFDSLLFNPIDVDISSKNEFLNHIKQKSVQYREILTTNESRFVALSTCEDMT
TDGRIIVIGQIE"

Spy0130 is referred to as a hypothetical protein. It contains a sortase substrate motif LPXTG (SEQ ID NO: 122), shown in italics in SEQ ID NO: 52.

SEQ ID NO: 52
MKKSILRILAIGYLLMSFCLLDSVEAENLTASINIEVINQVDVATNKQSSDIDETFMFVIEALDKESPLPNSVTT
SVKGNGKTSFEQLTFSEVGQYHYKIHQLLGKNSQYHYDETVYEVVIYVLYNEQSGALETNLVSNKLGETEKSELI
FKQEYSEKTPEPHQPDTTEKEKPQKKRNGI*LPSTG*EMVSYVSALGIVLVATITLYSIYKKLKTSK

Spy0130 contains an amino acid motif indicative of a cell wall anchor: **SEQ ID NO: 131** LPSTG (shown in italics in SEQ ID NO: 52, above). In some recombinant host cell systems, it may be preferable to remove this motif to facilitate secretion of a recombinant Spy0130 protein from the host cell. Alternatively, in other recombinant host cell systems, it may be preferable to use the cell wall anchor motif to anchor the recombinantly expressed protein to the cell wall. The extracellular domain of the expressed protein may be cleaved during purification or the recombinant protein may be left attached to either inactivated host cells or cell membranes in the final composition.

Two E boxes containing conserved glutamic residues have been identified in Spy0130. The E-box motifs are underlined in SEQ ID NO: 52, below. The conserved glutamic acid (E) residues, at amino acid residues 118 and 148, are marked in bold. The E box motifs, in particular the conserved glutamic acid residues, are thought to be important for the formation of oligomeric pilus-like structures of Spy0130. Preferred fragments of Spy0130 include at least one conserved glutamic acid residue. Preferably, fragments include at least one E box motif.

SEQ ID NO: 52
MKKSILRILAIGYLLMSFCLLDSVEAENLTASINIEVINQVDVATNKQSSDIDETFMFVIEALDKESPLPNSVTT
SVKGNGKTSFEQLTFSEVGQYHYKIHQLLGKNSQYHYDETVY**E**VVIYVLYNEQSGALETNLVSNKLGETEKS**E**LI
FKQEYSEKTPEPHQPDTTEKEKPQKKRNGILPSTGEMVSYVSALGIVLVATITLYSIYKKLKTSK

Spy0131 is referred to as a conserved hypothetical protein. An example of an amino acid sequence of Spy0131 is set forth in SEQ ID NO: 53

SEQ ID NO: 53
MTRTNYQKKRMTCPVETEDITYRRKKIKGRRQAILAQFEPELVHHELIGDSCTCPDCHGTLTEIGSVVQRQELVF
IPAQLKRINHVQHAYKCQTCSDNSLSDKIIKAPVPKAPLAHSLGSASIIAHTVHQKFTLKVPNYRQEEDWNKLGL
SISRKEIANWHIKSSQYYFEPLYDLLRDILLSQEVIHADETSYRVLESDTQLTYYWTFLSGKHEKKGITLYHHDK
RRSGLVTQEVLGDYSGYVHCDMHGAYRQLEHAKLVGCWAHVRRKFFEATPKQADKTSLGRKGLVYCDKLFALEAE
WCELPPQERLVKRKEILTPLMTTFFDWCREQVVLSGSKLGLAIAYSLKHERTFRTVLEDGHIVLSNNMAERAIKS
LVMGRKNWLFSQSFEGAKAAAIIMSLLETAKRHGLNSEKYISYLLDRLPNEETLAKREVLEAYLPWAKKVQTNCQ

Spy0133 is referred to as a conserved hypothetical protein. An example of an amino acid sequence of Spy0133 is set forth in SEQ ID NO: 54.

SEQ ID NO: 54
MTIRLNDLGQVYLVCGKTDMRQGIDSLAYLVKSQHELDLFSGAVYLFCGGRRDRFKALYWDGQGFWLLYKRFENG
KLAWPRNRDEVKCLTAVQVDWLMKGFFISPNIKISKSHDFY

Spy0135 is a SrtB type sortase. It is also referred to as a putative fibria-associated protein. An example of an amino acid sequence of Spy0135 is set forth in SEQ ID NO: 55.

SEQ ID NO: 55
MECYRDRQLLSTYHKQVTQKKPSEMEEVWQKAKAYNARLGIQPVPDAFSFRDGIHDKNYESLLQIENNDIMGYVE
VPSIKVTLPIYHYTTDEVLTKGAGHLFGSALPVGGDGTHTVISAHRGLPSAEMFTNLNLVKKGDTFYFRVLNKVL
AYKVDQILTVEPDQVTSLSGVMGKDYATLVTCTPYGVNTKRLLVRGHRIAYHYKKYQQAKKAMKLVDKSRMWAEV
VCAAFGVVIAIILVFMYSRVSAKKSK

GAS AI-3 sequences

As discussed above, a GAS AI-3 sequence is present in a M3, M18 and M5 strain isolates. Examples of GAS AI-3 sequences from M3 strain isolate MGAS315 are set forth below.

SpyM30097 is as a negative transcriptional regulator (Nra). An example of an amino acid sequence of SpyM30097 is set forth in SEQ ID NO: 56.

SEQ ID NO: 56
MPYVKKKKDSFLVETYLEQSIRDKSELVLLLFKSPTIIFSHVAKQTGLTAVQLKYYCKELDDFFGNNLDITIKKG
KIICCFVKPVKEFYLHQLYDTSTILKLLVFFIKNGTSSQPLIKFSKKYFLSSSSAYRLRESLIKLLREFGLRVSK
NTIVGEEYRIRYLIAMLYSKFGIVIYPLDHLDNQIIYRFLSQSATNLRTSPWLEEPFSFYNMLLALSWKRHQFAV
SIPQTRIFRQLKKLFIYDCLTRSSRQVIENAFSLTFSQGDLEYLFLIYITTNNSFASLQWTPQHIETCCHIFEKN
DTFRLLLEPILKRLPQLNHSKQDLIKALMYFSKSFLFNLQHFVIEIPSFSLPTYTGNSNLYKALKNIVNQWLAQL
PGKRHLNEKHLQLFCSHIEQILKNKQPALTVVLISSNFINAKLLTDTIPRYFSDKGIHFYSFYLLRDDIYQIPSL
KPDLVITHSRLIPFVKNDLVKGVTVAEFSFDNPDYSIASIQNLIYQLKDKKYQDFLNEQLQ

SpyM30098 is thought to be a collagen binding protein (Cpb). It contains a sortase substrate motif VPXTG (SEQ ID NO: 137) shown in italics in SEQ ID NO: 57.

SEQ ID NO: 57
MQKRDKTNYGSANNKRRQTTIGLLKVFLTFVALIGIVGFSIRAFGAEEQSVPNKQSSVQDYPWYGYDSYSKGYPD
YSPLKTYHNLKVNLDGSKEYQAYCFNLTKHFPSKSDSVRSQWYKKLEGTNENFIKLADKPRIEDGQLQQNILRIL
YNGYPNDRNGIMKGIDPLNAILVTQNAIWYYTDSSYISDTSKAFQQEETDLKLDSQQLQLMRNALKRLINPKEVE
SLPNQVPANYQLSIFQSSDKTFQNLLSAEYVPDTPPKPGEEPPAKTEKTSVIIRKYAEGDYSKLLEGATLKLAQI
EGSGFQEKIFDSNKSGEKVELPNGTYVLSELKPPQGYGVATPITFKVAAEKVLIKNKEGQFVENQNKEIAEPYSV
TAFNDFEEIGYLSDFNNYGKFYYAKNTNGTNQVVYCFNADLHSPPDSYDHGANIDPDVSESKEIKYTHVSGYDLY
KYAATPRDKDADFFLKHIKKILDKGYKKKGDTYKTLTEAQFRAATQLAIYYYTDSADLTTLKTYNDNKGYHGFDK
LDDATLAVVHELITYAEDVTLPMTQNLDFFVPNSSRYQALIGTQYHPNELIDVISMEDKQAPIIPITHKLTISKT
VTGTIADKKKEFNFEIHLKSSDGQAISGTYPTNSGELTVTDGKATFTLKDGESLIVEGLPSGYSYEITETGASDY
EVSVNGKNAPDGKATKASVKEDETVAFENRKDL*VPPTG*LTTDGAIYLWLLLLVPFGLLVWLFGRKGTKK

SpyM30098 contains an amino acid motif indicative of a cell wall anchor: **SEQ ID NO: 184** VPPTG (shown in italics in SEQ ID NO: 57, above). In some recombinant host cell systems, it may be preferable to remove this motif to facilitate secretion of a recombinant SpyM30098 protein from the host cell. Alternatively, in other recombinant host cell systems, it may be preferable to use the cell wall anchor motif to anchor the recombinantly expressed protein to the cell wall. The extracellular domain of the expressed protein may be cleaved during purification or the recombinant protein may be left attached to either inactivated host cells or cell membranes in the final composition.

A pilin motif, discussed above, containing a conserved lysine (K) residue has also been identified in SpyM30098. The pilin motif sequence is underlined in SEQ ID NO: 57, below. Conserved lysine (K) residues are also marked in bold, at amino acid residues 262 and 270. The pilin sequence, in particular the conserved lysine residues, are thought to be important for the formation of oligomeric, pilus-like structures. Preferred fragments of SpyM30098 include at least one conserved lysine residue. Preferably, fragments include the pilin sequence.

SEQ ID NO: 57
MQKRDKTNYGSANNKRRQTTIGLLKVFLTFVALIGIVGFSIRAFGAEEQSVPNKQSSVQDYPWYGYDSYSKGYPD
YSPLKTYHNLKVNLDGSKEYQAYCFNLTKHFPSKSDSVRSQWYKKLEGTNENFIKLADKPRIEDGQLQQNILRIL
YNGYPNDRNGIMKGIDPLNAILVTQNAIWYYTDSSYISDTSKAFQQEETDLKLDSQQLQLMRNALKRLINPKEVE
SLPNQVPANYQLSIFQSSDKTFQNLLSAEYVPDTPP**K**PGEEPPA**K**TEKTSVIIRKYAEGDYSKLLEGATLKLAQI
EGSGFQEKIFDSNKSGEKVELPNGTYVLSELKPPQGYGVATPITFKVAAEKVLIKNKEGQFVENQNKEIAEPYSV
TAFNDFEEIGYLSDFNNYGKFYYAKNTNGTNQVVYCFNADLHSPPDSYDHGANIDPDVSESKEIKYTHVSGYDLY
KYAATPRDKDADFFLKHIKKILDKGYKKKGDTYKTLTEAQFRAATQLAIYYYTDSADLTTLKTYNDNKGYHGFDK
LDDATLAVVHELITYAEDVTLPMTQNLDFFVPNSSRYQALIGTQYHPNELIDVISMEDKQAPIIPITHKLTISKT

VTGTIADKKKEFNFEIHLKSSDGQAISGTYPTNSGELTVTDGKATFTLKDGESLIVEGLPSGYSYEITETGASDY
EVSVNGKNAPDGKATKASVKEDETVAFENRKDLVPPTGLTTDGAIYLWLLLLVPFGLLVWLFGRKGTKK

An E box containing a conserved glutamic residue has been identified in SpyM30098. The E-box motif is underlined in SEQ ID NO: 57, below. The conserved glutamic acid (E), at amino acid residue 330, is marked in bold. The E box motif, in particular the conserved glutamic acid residue, is thought to be important for the formation of oligomeric pilus-like structures of SpyM30098. Preferred fragments of SpyM30098 include the conserved glutamic acid residue. Preferably, fragments include the E box motif.

SEQ ID NO: 57
MQKRDKTNYGSANNKRRQTTIGLLKVFLTFVALIGIVGFSIRAFGAEEQSVPNKQSSVQDYPWYGYDSYSKGYPD
YSPLKTYHNLKVNLDGSKEYQAYCFNLTKHFPSKSDSVRSQWYKKLEGTNENFIKLADKPRIEDGQLQQNILRIL
YNGYPNDRNGIMKGIDPLNAILVTQNAIWYYTDSSYISDTSKAFQQEETDLKLDSQQLQLMRNALKRLINPKEVE
SLPNQVPANYQLSIFQSSDKTFQNLLSAEYVPDTPPKPGEEPPAKTEKTSVIIRKYAEGDYSKLLEGATLKLAQI
EGSGFQEKIFDSNKSGEKVELPNGTYVLS**E**LKPPQGYGVATPITFKVAAEKVLIKNKEGQFVENQNKEIAEPYSV
TAFNDFEEIGYLSDFNNYGKFYYAKNTNGTNQVVYCFNADLHSPPDSYDHGANIDPDVSESKEIKYTHVSGYDLY
KYAATPRDKDADFFLKHIKKILDKGYKKKGDTYKTLTEAQFRAATQLAIYYYTDSADLTTLKTYNDNKGYHGFDK
LDDATLAVVHELITYAEDVTLPMTQNLDFFVPNSSRYQALIGTQYHPNELIDVISMEDKQAPIIPITHKLTISKT
VTGTIADKKKEFNFEIHLKSSDGQAISGTYPTNSGELTVTDGKATFTLKDGESLIVEGLPSGYSYEITETGASDY
EVSVNGKNAPDGKATKASVKEDETVAFENRKDLVPPTGLTTDGAIYLWLLLLVPFGLLVWLFGRKGTKK

SpyM30099 is referred to as LepA. An example of an amino acid sequence of SpyM30099 is set forth in SEQ ID NO: 58.

SEQ ID NO: 58
MTNYLNRLNENPLLKAFIRLVLKISIIGFLGYILFQYVFGVMIVNTNQMSPAVSAGDGVLYYRLTDRYHINDVVV
YEVDDTLKVGRIAAQAGDEVNFTQEGGLLINGHPPEKEVPYLTYPHSSGPNFPYKVPTGTYFILNDYREERLDSR
YYGALPINQIKGKISTLLRVRGI

SpyM30100 is thought to be a fimbrial protein. An example of an amino acid sequence of SpyM30100 is set forth in SEQ ID NO: 59.

SEQ ID NO: 59
MKKNKLLLATAILATALGTASLNQNVKAETAGVSENAKLIVKKTFDSYTDNEVLMPKADYTFKVEADSTASGKTK
DGLEIKPGIVNGLTEQIISYTNTDKPDSKVKSTEFDFSKVVFPGIGVYRYTVSEKQGDVEGITYDTKKWTVDVYV
GNKEGGGFEPKFIVSKEQGTDVKKPVNFNNSFATTSLKVKKNVSGNTGELQKEFDFTLTLNESTNFKKDQIVSLQ
KGNEKFEVKIGTPYKFKLKNGESIQLDKLPVGITYKVNEMEANKDGYKTTASLKEGDGQSKMYQLDMEQKTDESA
DEIVVTNKRDT*QVPTG*VVGTLAPFAVLSIVAIGGVIYITKRKKA

SpyM30100 contains an amino acid motif indicative of a cell wall anchor: **SEQ ID NO: 140** QVPTG (shown in italics in SEQ ID NO: 59, above). In some recombinant host cell systems, it may be preferable to remove this motif to facilitate secretion of a recombinant SpyM30100 protein from the host cell. Alternatively, in other recombinant host cell systems, it may be preferable to use the cell wall anchor motif to anchor the recombinantly expressed protein to the cell wall. The extracellular domain of the expressed protein may be cleaved during purification or the recombinant protein may be left attached to either inactivated host cells or cell membranes in the final composition.

Two pilin motifs, discussed above, containing conserved lysine (K) residues have also been identified in SpyM30100. The pilin motif sequences are underlined in SEQ ID NO: 59, below. Conserved lysine (K) residues are also marked in bold, at amino acid residues 57 and 63 and at amino acid residues 161 and 166. The pilin sequences, in particular the conserved lysine residues, are thought to be important for the formation of oligomeric, pilus-like structures. Preferred fragments of SpyM30100 include at least one conserved lysine residue. Preferably, fragments include at least one pilin sequence.

SEQ ID NO: 59
MKKNKLLLATAILATALGTASLNQNVKAETAGVSENAKLIVKKTFDSYTDNEVLMP**K**ADYTF**K**VEADSTASGKTK
DGLEIKPGIVNGLTEQIISYTNTDKPDSKVKSTEFDFSKVVFPGIGVYRYTVSEKQGDVEGITYDTKKWTVDVYV
GNKEGGGFEP**K**FIVS**K**EQGTDVKKPVNFNNSFATTSLKVKKNVSGNTGELQKEFDFTLTLNESTNFKKDQIVSLQ
KGNEKFEVKIGTPYKFKLKNGESIQLDKLPVGITYKVNEMEANKDGYKTTASLKEGDGQSKMYQLDMEQKTDESA
DEIVVTNKRDTQVPTGVVGTLAPFAVLSIVAIGGVIYITKRKKA

Two E boxes, each containing a conserved glutamic residue, have been identified in SpyM30100. The E-box motifs are underlined in SEQ ID NO: 59, below. The conserved glutamic acid (E) residues, at amino acid residues 232 and 264, are marked in bold. The E box motifs, in particular the conserved glutamic acid residues, are thought to be important for the formation of oligomeric pilus-like structures of SpyM30100. Preferred fragments of SpyM30100 include at least one conserved glutamic acid residue. Preferably, fragments include at least one E box motif.

SEQ ID NO: 59
MKKNKLLLATAILATALGTASLNQNVKAETAGVSENAKLIVKKTFDSYTDNEVLMPKADYTFKVEADSTASGKTK
DGLEIKPGIVNGLTEQIISYTNTDKPDSKVKSTEFDFSKVVFPGIGVYRYTVSEKQGDVEGITYDTKKWTVDVYV
GNKEGGGFEPKFIVSKEQGTDVKKPVNFNNSFATTSLKVKKNVSGNTGELQKEFDFTLTLNESTNFKKDQIVSLQ
KGNEKF**E**VKIGTPYKFKLKNGESIQLDKLPVGITYKVN**E**MEANKDGYKTTASLKEGDGQSKMYQLDMEQKTDESA
DEIVVTNKRDTQVPTGVVGTLAPFAVLSIVAIGGVIYITKRKKA

SpyM30101 is a SrtC2 type sortase. An example of an amino acid sequence of SpyM30101 is set forth in SEQ ID NO: 60.

SEQ ID NO: 60
MTIVQVINKAIDTLILIFCLVVLFLAGFGLWDSYHLYQQADASNFKKFKTAQQQPKFEDLLALNEDVIGWLNIPG
THIDYPLVQGKTNLEYINKAVDGSVAMSGSLFLDTRNHNDFTDDYSLIYGHHMAGNAMFGEIPKFLKKDFFSKHN
KAIIETKERKKLTVTIFACLKTDAFNQLVFNPNAITNQDQQRQLVDYISKRSKQFKPVKLKHHTKFVAFSTCENF
STDNRVIVVGTIQE

SpyM30102 is referred to as a hypothetical protein. An example of an amino acid sequence of SpyM30102 is set forth in SEQ ID NO: 61.

SEQ ID NO: 61
MILTMLAFNQTVLAKDSTVQTSISVENVLERAGDSTPFSIALESIDAMKTIEEITIAGSGKASFSPLTFTTVGQY
TYRVYQKPSQNKDYQADTTVFDVLVYVTYDEDGTLVAKVISRRAGDEEKSAITFKPKWLVKPIPPRQPNIPKTP*L*
*PLAG*EVKSLLGILSIVLLGLLVLLYVKKLKSRL

SpyM30102 contains an amino acid motif indicative of a cell wall anchor: **SEQ ID NO: 185** LPLAG (shown in italics in SEQ ID NO: 61, above). In some recombinant host cell systems, it may be preferable to remove this motif to facilitate secretion of a recombinant SpyM30102 protein from the host cell. Alternatively, in other recombinant host cell systems, it may be preferable to use the cell wall anchor motif to anchor the recombinantly expressed protein to the cell wall. The extracellular domain of the expressed protein may be cleaved during purification or the recombinant protein may be left attached to either inactivated host cells or cell membranes in the final composition.

A pilin motif, discussed above, containing a conserved lysine (K) residue has also been identified in SpyM30102. The pilin motif sequence is underlined in SEQ ID NO: 61, below. The conserved lysine (K) residue is also marked in bold, at amino acid residue 132. The pilin sequence, in particular the conserved lysine residues, are thought to be important for the formation of oligomeric, pilus-like structures. Preferred fragments of SpyM30102 include the conserved lysine residue. Preferably, fragments include the pilin sequence.

SEQ ID NO: 61
MILTMLAFNQTVLAKDSTVQTSISVENVLERAGDSTPFSIALESIDAMKTIEEITIAGSGKASFSPLTFTTVGQY
TYRVYQKPSQNKDYQADTTVFDVLVYVTYDEDGTLVAKVISRRAGDEEKSAITFKPKWLVKPIPPRQPNIPKTPL
PLAGEVKSLLGILSIVLLGLLVLLYVKKLKSRL

Two E boxes containing conserved glutamic residues have been identified in SpyM30102. The E-box motifs are underlined in SEQ ID NO: 61, below. The conserved glutamic acid (E) residues, at amino acid residues 52 and 122, are marked in bold. The E box motifs, in particular the conserved glutamic acid residues, are thought to be important for the formation of oligomeric pilus-like structures of SpyM30102. Preferred fragments of SpyM30102 include at least one conserved lysine residue. Preferably, fragments include at least one pilin sequence.

SEQ ID NO: 61
MILTMLAFNQTVLAKDSTVQTSISVENVLERAGDSTPFSIALESIDAMKTI**E**EITIAGSGKASFSPLTFTTVGQY
TYRVYQKPSQNKDYQADTTVFDVLVYVTYDEDGTLVAKVISRRAGD**E**EKSAITFKPKWLVKPIPPRQPNIPKTPL
PLAGEVKSLLGILSIVLLGLLVLLYVKKLKSRL

SpyM30103 is referred to as a putative multiple sugar metabolism regulator. An example of an amino acid sequence for SpyM3103 is set forth in SEQ ID NO: 62.

SEQ ID NO: 62
MVRFDLKHVQTLHSLSQLPISVMSQDKALIQVYGNDDYLLCYYQFLKHLAIPQAAQDVIFYEGLFEESFMIFPLC
HYIIAIGPFYPYSLNKDYQEQLANNCLKHSSHRSKEELLSYMALVPHFPINNVRNLLIAIDAFFDTQFETTCQQT
IHQLLQHSKQMTADPDIIHRLKHISKASSQLPPVLEHLNHIMDLVKLGNPQLLKQEINRIPLSSITSSSISALRA
EKNLTVIYLTRLLEFSFVENTDVAKHYSLVKYYMALNEEASDLLKVLRIRCAAIIHFSESLTNKSISDKRQMYNS
VLHYVDSHLYSKLKVSDIAKRLYVSESHLRSVFKKYSNVSLQHYILSTKIKEAQLLLKRGIPVGEVAKSLYFYDT
THFHKIFKKYTGISSKDYLAKYRDNI

SpyM30104 is thought to be a F2 like fibronectic binding protein. An example of an amino acid sequence for SpyM30104 is set forth in SEQ ID NO: 63.

SEQ ID NO: 63
MSSSDEETLKQYASKYTSNRRGDTSGNLKKQIAKVLTEGYPTNKSDWLNGLTENEKIEVTQDAIWYFTETTVPAD
RSYTNRNVNSQKMKEVYQKLIDTTDIDKYEDVQFDLFVPQDTNLQAVISVEPVIESLPWTSLKPIAQKDITAKKI
WVDAPKEKPIIYFKLYRQLPGEKEVAVDDAELKQINSEGQQEISVTWTNQLVTDEKGMAYIYSVKEVDKNGELLE
PKDYIKKEDGLTVTNTYVKPTSGHYDIEVTFGNGHIDITEDTTPDIVSGENQMKQIEGEDSKPIDEVTENNLIEF
GKNTMPGEEDGTNSNKYEEVEDSRPVDTLSGLSSEQGQSGDMTIEEDSATHIKFSKRDIDGKELAGATMELRDSS
GKTISTWISDGQVKDFYLMPGKYTFVETAAPDGYEVATAITFTVNEQGQVTVNGKATKGDAHIVMVDAYKPTKGS
GQVIDIEEKLPDEQGHSGSTTEIEDSKSSDVIIGGQGEVVDTTEDTQSGMTGHSGSTTEIEDSKSSDVIIGGQGE
VVDTTEDTQSGMTGHSGSTTKIEDSKSSDVIVGGQGQIVETTEDTQTGMHGDSGRKTEVEDTKLVQSFHFDNKEP
ESNSEIPKKDKSKSNTS*LPATG*EKQHNKFFWMVTSCSLISSVFVISLKSKKRLSSC

SpyM30104 contains an amino acid motif indicative of a cell wall anchor: **SEQ ID NO: 180** LPATG (shown in italics in SEQ ID NO: 63, above). In some recombinant host cell systems, it may be preferable to remove this motif to facilitate secretion of a recombinant SpyM30104 protein from the host cell. Alternatively, in other recombinant host cell systems, it may be preferable to use the cell wall anchor motif to anchor the recombinantly expressed protein to the cell wall. The extracellular domain of the expressed protein may be cleaved during purification or the recombinant protein may be left attached to either inactivated host cells or cell membranes in the final composition.

Two pilin motifs, discussed above, containing conserved lysine (K) residues have also been identified in SpyM30104. The pilin motif sequences are underlined in SEQ ID NO: 63, below. Conserved lysine (K) residues are also marked in bold, at amino acid residues 156 and 227. The pilin sequences, in particular the conserved lysine residues, are thought to be important for the formation of oligomeric, pilus-like structures. Preferred fragments of SpyM30104 include at least one conserved lysine residue. Preferably, fragments include at least one pilin sequence.

SEQ ID NO: 63
MSSSDEETLKQYASKYTSNRRGDTSGNLKKQIAKVLTEGYPTNKSDWLNGLTENEKIEVTQDAIWYFTETTVPAD
RSYTNRNVNSQKMKEVYQKLIDTTDIDKYEDVQFDLFVPQDTNLQAVISVEPVIESLPWTSLKPIAQKDITAKKI
WVDAP**K**EKPIIYFKLYRQLPGEKEVAVDDAELKQINSEGQQEISVTWTNQLVTDEKGMAYIYSVKEVDKNGELLE
P**K**DYIKKEDGLTVTNTYVKPTSGHYDIEVTFGNGHIDITEDTTPDIVSGENQMKQIEGEDSKPIDEVTENNLIEF
GKNTMPGEEDGTNSNKYEEVEDSRPVDTLSGLSSEQGQSGDMTIEEDSATHIKFSKRDIDGKELAGATMELRDSS
GKTISTWISDGQVKDFYLMPGKYTFVETAAPDGYEVATAITFTVNEQGQVTVNGKATKGDAHIVMVDAYKPTKGS
GQVIDIEEKLPDEQGHSGSTTEIEDSKSSDVIIGGQGEVVDTTEDTQSGMTGHSGSTTEIEDSKSSDVIIGGQGE
VVDTTEDTQSGMTGHSGSTTKIEDSKSSDVIVGGQGQIVETTEDTQTGMHGDSGRKTEVEDTKLVQSFHFDNKEP
ESNSEIPKKDKSKSNTSLPATGEKQHNKFFWMVTSCSLISSVFVISLKSKKRLSSC

An E box containing a conserved glutamic residue has been identified in SpyM30104. The E-box motif is underlined in SEQ ID NO: 63, below. The conserved glutamic acid (E), at amino acid residue 402, is marked in bold. The E box motif, in particular the conserved glutamic acid residue, is thought to be important for the formation of oligomeric pilus-like structures of SpyM30104. Preferred fragments of SpyM30104 include the conserved glutamic acid residue. Preferably, fragments include the E box motif.

SEQ ID NO: 63
MSSSDEETLKQYASKYTSNRRGDTSGNLKKQIAKVLTEGYPTNKSDWLNGLTENEKIEVTQDAIWYFTETTVPAD
RSYTNRNVNSQKMKEVYQKLIDTTDIDKYEDVQFDLFVPQDTNLQAVISVEPVIESLPWTSLKPIAQKDITAKKI
WVDAPKEKPIIYFKLYRQLPGEKEVAVDDAELKQINSEGQQEISVTWTNQLVTDEKGMAYIYSVKEVDKNGELLE
PKDYIKKEDGLTVTNTYVKPTSGHYDIEVTFGNGHIDITEDTTPDIVSGENQMKQIEGEDSKPIDEVTENNLIEF
GKNTMPGEEDGTNSNKYEEVEDSRPVDTLSGLSSEQGQSGDMTIEEDSATHIKFSKRDIDGKELAGATMELRDSS
GKTISTWISDGQVKDFYLMPGKYTFV**E**TAAPDGYEVATAITFTVNEQGQVTVNGKATKGDAHIVMVDAYKPTKGS
GQVIDIEEKLPDEQGHSGSTTEIEDSKSSDVIIGGQGEVVDTTEDTQSGMTGHSGSTTEIEDSKSSDVIIGGQGE
VVDTTEDTQSGMTGHSGSTTKIEDSKSSDVIVGGQGQIVETTEDTQTGMHGDSGRKTEVEDTKLVQSFHFDNKEP
ESNSEIPKKDKSKSNTSLPATGEKQHNKFFWMVTSCSLISSVFVISLKSKKRLSSC

Examples of GAS AI-3 sequences from M3 strain isolate SSI-1 are set forth below.

Sps0099 is a negative transcriptional regulator (Nra). An example of an amino acid sequence for Sps0099 is set forth in SEQ ID NO: 64.

SEQ ID NO: 64
MPYVKKKKDSFLVETYLEQSIRDKSELVLLLFKSPTIIFSHVAKQTGLTAVQLKYYCKELDDFFGNNLDITIKKG
KIICCFVKPVKEFYLHQLYDTSTILKLLVFFIKNGTSSQPLIKFSKKYFLSSSSAYRLRESLIKLLREFGLRVSK
NTIVGEEYRIRYLIAMLYSKFGIVIYPLDHLDNQIIYRFLSQSATNLRTSPWLEEPFSFYNMLLALSWKRHQFAV
SIPQTRIFRQLKKLFIYDCLTRSSRQVIENAFSLTFSQGDLEYLFLIYITTNNSFASLQWTPQHIETCCHIFEKN
DTFRLLLEPILKRLPQLNHSKQDLIKALMYFSKSFLFNLQHFVIEIPSFSLPTYTGNSNLYKALKNIVNQWLAQL
PGKRHLNEKHLQLFCSHIEQILKNKQPALTVVLISSNFINAKLLTDTIPRYFSDKGIHFYSFYLLRDDIYQIPSL
KPDLVITHSRLIPFVKNDLVKGVTVAEFSFDNPDYSIASIQNLIYQLKDKKYQDFLNEQLQ

Sps0100 is thought to be a collagen binding protein (Cbp). It contains a sortase substrate motif VPXTG shown in italics in SEQ ID NO: 65.

SEQ ID NO: 65

MQKRDKTNYGSANNKRRQTTIGLLKVELTIVALIGIVGFSIRAFGAEEQSVPNKQSSVQDYPWYGYDSYSKGYPD
YSPLKTYHNLKVNLDGSKEYQAYCFNLTKHFPSKSDSVRSQWYKKLEGTNENFIKLADKPRIEDGQLQQNILRIL
YNGYPNDRNGIMKGIDPLNAILVTQNAIWYYTDSSYISDTSKAFQQEETDLKLDSQQLQLMRNALKRLINPKEVE
SLPNQVPANYQLSIFQSSDKTFQNLLSAEYVPDTPPKPGEEPPAKTEKTSVIIRKYAEGDYSKLLEGATLKLAQI
EGSGFQEKIFDSNKSGEKVELPNGTYVLSELKPPQGYGVATPITFKVAAEKVLIKNKEGQFVENQNKEIAEPYSV
TAFNDFEEIGYLSDFNNYGKFYYAKNTNGTNQVVYCFNADLHSPPDSYDHGANIDPDVSESKEIKYTHVSGYDLY
KYAATPRDKDADFFLKHIKKILDKGYKKKGDTYKTLTEAQFRAATQLAIYYYTDSADLTTLKTYNDNKGYHGFDK
LDDATLAVVHELITYAEDVTLPMTQNLDFFVPNSSRYQALIGTQYHPNELIDVISMEDKQAPIIPITHKLTISKT
VTGTIADKKKEFNFEIHLKSSDGQAISGTYPTNSGELTVTDGKATFTLKDGESLIVEGLPSGYSYEITETGASDY
EVSVNGKNAPDGKATKASVKEDETVAFENRKDL*VPPTG*LTTDGAIYLWLLLLVPFGLLVWLFGRKGTKK

Sps0101 is referred to as a LepA protein. An example of an amino acid sequence of Sps0101 is set forth as SEQ ID NO: 66

SEQ ID NO: 66
MTNYLNRLNENPLLKAFIRLVLKISIIGFLGYILFQYVFGVMIVNTNQMSPAVSAGDGVLYYRLTDRYHINDVVV
YEVDDTLKVGRIAAQAGDEVNFTQEGGLLINGHPPEKEVPYLTYPHSSGPNFPYKVPTGTYFILNDYREERLDSR
YYGALPINQIKGKISTLLRVRGI

Sps0102 is thought to be a fimbrial protein. It contains a sortase substrate motif QVXTG shown in italics in SEQ ID NO: 67.

SEQ ID NO: 67
MEREKMKKNKLLLATAILATALGTASLNQNVKAETAGVSENAKLIVKKTFDSYTDNEVLMPKADYTFKVEADSTA
SGKTKDGLEIKPGIVNGLTEQIISYTNTDKPDSKVKSTEFDFSKVVFPGIGVYRYTVSEKQGDVEGITYDTKKWT
VDVYVGNKEGGGFEPKFIVSKEQGTDVKKPVNFNNSFATTSLKVKKNVSGNTGELQKEFDFTLTLNESTNFKKDQ
IVSLQKGNEKFEVKIGTPYKFKLKNGESIQLDKLPVGITYKVNEMEANKDGYKTTASLKEGDGQSKMYQLDMEQK
TDESADEIVVTNKRDT*QVPTG*VVGTLAPFAVLSIVAIGGVIYITKRKKA

Sps0103 is a SrtC2 type sortase. An example of Sps0103 is set forth in SEQ ID NO: 68.

SEQ ID NO: 68
MVMTIVQVINKAIDTLILIFCLVVLFLAGFGLWDSYHLYQQADASNFKKFKTAQQQPKFEDLLALNEDVIGWLNI
PGTHIDYPLVQGKTNLEYINKAVDGSVAMSGSLFLDTRNHNDFTDDYSLIYGHHMAGNAMFGEIPKFLKKDFFSK
HNKAIIETKERKKLTVTIFACLKTDAFNQLVFNPNAITNQDQQRQLVDYISKRSKQFKPVKLKHHTKFVAFSTCE
NFSTDNRVIVVGTIQE

Sps0104 is referred to as a hypothetical protein. It contains a sortase substrate motif LPXAG shown in italics in SEQ ID NO: 69.

SEQ ID NO: 69
MLFSVVMILTMLAFNQTVLAKDSTVQTSISVENVLERAGDSTPFSIALESIDAMKTIEEITIAGSGKASFSPLTF
TTVGQYTYRVYQKPSQNKDYQADTTVFDVLVYVTYDEDGTLVAKVISRRAGDEEKSAITFKPKWLVKPIPPRQPN
IPKTP*LPLAG*EVKSLLGILSIVLLGLLVLLYVKKLKSRL

Sps0105 is referred to as a putative multiple sugar metabolism regulator. An example of Sps0105 is set forth in SEQ ID NO: 70.

SEQ ID NO: 70
MALVPHFPINNVRNLLIAIDAFFDTQFETTCQQTIHQLLQHSKQMTADPDIIHRLKHISKASSQLPPVLEHLNHI
MDLVKLGNPQLLKQEINRIPLSSITSSSISALRAEKNLTVIYLTRLLEFSFVENTDVAKHYSLVKYYMALNEEAS
DLLKVLRIRCAAIIHFSESLTNKSISDKRQMYNSVLHYVDSHLYSKLKVSDIAKRLYVSESHLRSVFKKYSNVSL
QHYILSTKIKEAQLLLKRGIPVGEVAKSLYFYDTTHFHKIFKKYTGISSKDYLAKYRDNI

Sps0106 is thought to be a F2 like fibronectic binding protein. It contains a sortase substrate LPXTG (SEQ ID NO: 122) shown in italics in SEQ ID NO: 71.

SEQ ID NO: 71

MTQKNSYKLSFLLSLTGEILGLLLVFIGLSGVSVGHAETRNGANKQGAFEIKKNKSQEEYNYEVYDNRNILQDGE
HKLEIKRVDGTGKTYQGFCFQLTKNFPTAQGVSKKLYKKLSSSDEETLKQYASKYTSNRRGDTSGNLKKQIAKVL
TEGYPTNKSDWLNGLTENEKIEVTQDAIWYFTETTVPADRSYTNRNVNSQKMKEVYQKLIDTTDIDKYEDVQFDL
FVPQDTNLQAVISVEPVIESLPWTSLKPIAQKDITAKKIWVDAPKEKPIIYFKLYRQLPGEKEVAVDDAELKQIN
SEGQQEISVTWTNQLVTDEKGMAYIYSVKEVDKNGELLEPKDYIKKEDGLTVTNTYVKPTSGHYDIEVTFGNGHI
DITEDTTPDIVSGENQMKQIEGEDSKPIDEVTENNLIEFGKNTMPGEEDGTNSNKYEEVEDSRPVDTLSGLSSEQ
GQSGDMTIEEDSATHIKFSKRDIDGKELAGATMELRDSSGKTISTWISDGQVKDFYLMPGKYTFVETAAPDGYEV
ATAITFTVNEQGQVTVNGKATKGDAHIVMVDAYKPTKGSGQVIDIEEKLPDEQGHSGSTTEIEDSKSSDVIIGGQ
GEVVDTTEDTQSGMTGHSGSTTKIEDSKSSDVIVGGQGQIVETTEDTQTGMHGDSGRKTEVEDTKLVQSFHFDNK
EPESNSEIPKKDKSKSNTS*LPATG*EKQHNKFFWMVTSCSLISSVFVISLKSKKRLSSC

Examples of GAS AI-3 sequences from M5 isolate Manfredo are set forth below.

Orf 77 encodes a negative transcription regulator (Nra). An example of the nucleotide sequence encoding Nra (SEQ ID NO: 88) and an Nra amino acid sequence (SEQ ID NO: 89) are set forth below.

SEQ ID NO: 88

ATGCCTTATGTCAAAAAGAAAAAGGATAGTTTCTTAGTAGAAACATATCTTGAACAGTCTATTAGAGATAAAAGT
GAATTAGTCTTACTGTTATTTAAATCGCCTACTATCATTTTTTCTCATGTTGCTAAACAAACTGGTCTGACGGCT
GTACAATTAAAATATTACTGTAAAGAACTTGATGACTTTTTTGGAAATAATTTAGACATTACCATTAAAAAGGGC
AAAATAATATGTTGTTTTGTCAAACCTGTTAAGGAATTCTACCTTCATCAACTCTATGACACATCAACAATATTA
AAATTATTAGTTTTCTTTATTAAAAATGGAACGTCATCACAACCTCTGATTAAATTTTCAAAAAAGTATTTTCTA
TCAAGCTCCTCAGCTTATCGACTACGGGAATCGCTGATCAAATTACTACGGGAATTTGGCTTGAGAGTCTCAAAA
AATACAATTGTCGGAGAGGAATATCGTATTCGCTATCTTATTGCCATGCTATATAGTAAATTTGGCATTGTCATC
TATCCGTTAGATCATCTAGACAATCAAATTATTTATCGCTTCTTATCACAAAGTGCAACCAATTTAAGAACATCG
CCCTGGCTAGAGGAACCTTTTTCTTTTTATAATATGTTACTTGCCTTGTCATGGAAACGTCACCAATTTGCAGTT
AGCATTCCTCAAACACGTATTTTTCGACAATTAAAAAAGCTTTTTATCTATGATTGTTTAACTCGAAGCAGTCGA
CAAGTAATCGAAAATGCTTTTTCGTTAATGTTCTCACAAGGAGATCTCGATTATCTTTTTTTAATTTATATTACC
ACCAATAATTCCTTTGCCAGCCTACAATGGACTCCACAGCATATTGAAACTTGCTGCCATATTTTTGAAAAAAAT
GACACATTTCGGTTATTGTTAGAGCCCATTCTTAAACGTTTACCGCAATTAAACCATTCTAAACAAGACCTTATT
AAAGCCCTTATGTATTTTTCAAAATCTTTTCTATTTAACCTCCAACATTTCGTCATCGAGATTCCTTCTTTTTCC
TTGCCGACCTATACAGGCAACTCTAATCTTTACAAAGCTTTAAAAAATATTGTAAATCAGTGGCTTGCTCAATTA
CCCGGAAAGCGTCATCTTAACGAAAAGCATCTCCAACTTTTTTGCTCTCATATTGAACAAATCTTAAAAAATAAA
CAACCTGCTTTAACTGTCGTTTTAATATCTAGTAACTTTATAAATGCTAAACTCCTTACAGATACTATCCCACGA
TATTTTTCTGATAAAGGAATTCATTTTTATTCTTTTTACTTATTAAGAGATGATATCTATCAAATTCCAAGCTTA
AAACCAGATTTAGTTATCACTCATAGCCGATTAATTCCTTTTGTTAAGAATGATCTGGTCAAAGGTGTTACTGTT
GCTGAATTTTCTTTTGATAACCCTGACTACTCTATTGCTTCAATTCAAAACTTGATATATCAGCTCAAAGATAAA
AAATATCAAGATTTTCTAAACGAGCAATTACAA

SEQ ID NO: 89

MPYVKKKKDSFLVETYLEQSIRDKSELVLLLFKSPTIIFSHVAKQTGLTAVQLKYYCKELDDFFGNNLDITIKKG
KIICCFVKPVKEFYLHQLYDTSTILKLLVFFIKNGTSSQPLIKFSKKYFLSSSSAYRLRESLIKLLREFGLRVSK
NTIVGEEYRIRYLIAMLYSKFGIVIYPLDHLDNQIIYRFLSQSATNLRTSPWLEEPFSFYNMLLALSWKRHQFAV
SIPQTRIFRQLKKLFIYDCLTRSSRQVIENAFSLMFSQGDLDYLFLIYITTNNSFASLQWTPQHIETCCHIFEKN
DTFRLLLEPILKRLPQLNHSKQDLIKALMYFSKSFLFNLQHFVIEIPSFSLPTYTGNSNLYKALKNIVNQWLAQL
PGKRHLNEKHLQLFCSHIEQILKNKQPALTVVLISSNFINAKLLTDTIPRYFSDKGIHFYSFYLLRDDIYQIPSL
KPDLVITHSRLIPFVKNDLVKGVTVAEFSFDNPDYSIASIQNLIYQLKDKKYQDFLNEQLQ

Orf 78 is thought to be a collagen binding protein (Cbp). An example of the nucleotide sequence encoding Cbp (SEQ ID NO: 90) and a Cbp amino acid sequence (SEQ ID NO: 91) are set forth below.

SEQ ID NO: 90

TTGCAAAAGAGGGATAAAACCAATTATGGAAGCGCTAACAACAAACGACGACAAACGACGATCGGATTACTGAAA
GTATTTTTGACGTTTGTAGCTCTGATAGGAATAGTAGGGTTTTCTATCAGAGCGTTCGGAGCTGAAGAAAAATCT
ACTGAAACTAAAAAAACGTCAGTCATTATTAGAAAATATGCTGAAGGTGACTACTCTAAACTTCTAGAGGGAGCA
ACTTTGCGTTTAACAGGGGAAGATATCCCAGATTTTCAAGAAAAAGTCTTCCAAAGTAATGGAACAGGAGAAAAG
ATTGAATTATCAAATGGGACTTATACCTTAACAGAAACATCATCTCCAGATGGATATAAAATTACGGAGCCGATT

AAGTTTAGAGTAGTGAATAAAAAAGTATTTATCGTCCAAAAAGATGGTTCTCAAGTGGAAAACCCAAACAAAGAA
CTAGGTTCTCCATATACTATAGAGGCATACAATGATTTTGATGAATTTGGCTTACTGTCAACACAAAATTATGCG
AAATTTTATTATGGAAAAAACTATGATGGCAGTTCACAAATTGTTTATTGCTTCAATGCCAACTTGAAATCTCCA
CCTGACTCGGAAGATCATGGTGCTACAATAAATCCTGACTTTACGACTGGTGATATTAGGTACAGTCATATTGCT
GGTTCAGATTTGATAAAATACGCTAATACAGCTAGGGATGAAGATCCTCAATTATTTTTAAAACACGTAAAAAAA
GTAATTGAAAATGGGTATCATAAAAAAGGTCAAGCTATTCCATATAACGGTCTGACTGAGGCACAGTTTCGTGCG
GCTACTCAACTGGCAATTTATTATTTTACAGATAGTGTTGACTTAACTAAGGATAGATTGAAAGACTTCCATGGA
TTTGGAGATATGAATGATCAAACTTTGGGTGTAGCTAAAAAAATTGTAGAATACGCTTTGAGTGATGAAGATTCA
AAACTAACAAATCTTGATTTCTTCGTACCTAATAATAGCAAATACCAATCTCTTATTGGGACAGAATACCATCCA
GATGATTTGGTTGACGTGATTCGTATGGAAGATAAAAAGCAAGAAGTTATTCCAGTAACTCATAGTTTGACGGTG
CAAAAAACAGTAGTCGGTGAGTTGGGAGATAAGACTAAAGGCTTTCAATTTGAACTTGAGTTGAAAGATAAAACT
GGACAGCCTATTGTTAACACTCTAAAAACTAATAATCAAGATTTAGTAGCTAAAGATGGGAAATATTCATTTAAT
CTAAAGCATGGTGACACCATAAGAATAGAAGGATTACCGACGGGATATTCTTATACCCTGAAAGAGACTGAAGCT
AAGGATTATATAGTAACTGTTGATAACAAAGTTAGTCAAGAAGCTCAATCAGCAAGTGAGAATGTCACAGCAGAC
AAAGAAGTCACTTTTGAAAACCGAAAAGATCTTGTCCCACCAACTGGTTTGACAACAGATGGGGCTATCTATCTT
TGGTTATTACTACTTGTTCCATTTGGGTTATTGGTTTGGCTATTTGGTCGTAAAGGGTTAAAAAATGAC

SEQ ID NO: 91

MQKRDKTNYGSANNKRRQTTIGLLKVFLTFVALIGIVGFSIRAFGAEEKSTETKKTSVIIRKYAEGDYSKLLEGA
TLRLTGEDIPDFQEKVFQSNGTGEKIELSNGTYTLTETSSPDGYKITEPIKFRVVNKKVFIVQKDGSQVENPNKE
LGSPYTIEAYNDFDEFGLLSTQNYAKFYYGKNYDGSSQIVYCFNANLKSPPDSEDHGATINPDFTTGDIRYSHIA
GSDLIKYANTARDEDPQLFLKHVKKVIENGYHKKGQAIPYNGLTEAQFRAATQLAIYYFTDSVDLTKDRLKDFHG
FGDMNDQTLGVAKKIVEYALSDEDSKLTNLDFFVPNNSKYQSLIGTEYHPDDLVDVIRMEDKKQEVIPVTHSLTV
QKTVVGELGDKTKGFQFELELKDKTGQPIVNTLKTNNQDLVAKDGKYSFNLKHGDTIRIEGLPTGYSYTLKETEA
KDYIVTVDNKVSQEAQSASENVTADKEVTFENRKDL*VPPTG*LTTDGAIYLWLLLLVPFGLLVWLFGRKGLKND

Orf 78 contains an amino acid motif indicative of a cell wall anchor: **SEQ ID NO: 184** VPPTG (shown in italics in SEQ ID NO: 91, above). In some recombinant host cell systems, it may be preferable to remove this motif to facilitate secretion of a recombinant Orf 78 protein from the host cell. Alternatively, in other recombinant host cell systems, it may be preferable to use the cell wall anchor motif to anchor the recombinantly expressed protein to the cell wall. The extracellular domain of the expressed protein may be cleaved during purification or the recombinant protein may be left attached to either inactivated host cells or cell membranes in the final composition.

Three E boxes containing conserved glutamic residues have been identified in Orf 78. The E-box motifs are underlined in SEQ ID NO: 91, below. The conserved glutamic acid (E) residues, at amino acid residues 112, 395, and 447, are marked in bold. The E box motifs, in particular the conserved glutamic acid residues, are thought to be important for the formation of oligomeric pilus-like structures of Orf 78. Preferred fragments of Orf 78 include at least one conserved glutamic acid residue. Preferably, fragments include at least one E box motif.

SEQ ID NO: 91

MQKRDKTNYGSANNKRRQTTIGLLKVFLTFVALIGIVGFSIRAFGAEEKSTETKKTSVIIRKYAEGDYSKLLEGA
TLRLTGEDIPDFQEKVFQSNGTGEKIELSNGTYTLT**E**TSSPDGYKITEPIKFRVVNKKVFIVQKDGSQVENPNKE
LGSPYTIEAYNDFDEFGLLSTQNYAKFYYGKNYDGSSQIVYCFNANLKSPPDSEDHGATINPDFTTGDIRYSHIA
GSDLIKYANTARDEDPQLFLKHVKKVIENGYHKKGQAIPYNGLTEAQFRAATQLAIYYFTDSVDLTKDRLKDFHG
FGDMNDQTLGVAKKIVEYALSDEDSKLTNLDFFVPNNSKYQSLIGTEYHPDDLVDVIRMEDKKQEVIPVTHSLTV
QKTVVGELGDKTKGFQFEL**E**LKDKTGQPIVNTLKTNNQDLVAKDGKYSFNLKHGDTIRIEGLPTGYSYTLK**E**TEA
KDYIVTVDNKVSQEAQSASENVTADKEVTFENRKDLVPPTGLTTDGAIYLWLLLLVPFGLLVWLFGRKGLKND

Orf 79 is thought to be a LepA signal peptidase I. An example of the nucleotide sequence encoding a LepA signal peptidase I (SEQ ID NO: 92) and a LepA signal peptidase I amino acid sequence (SEQ ID NO: 93) are set forth below.

SEQ ID NO: 92

ATGACTAATTACCTAAATCGTTTAAATGAGAATTCACTATTTAAAGCTTTCATACGGTTAGTACTTAAGATTTCT
ATTATTGGGTTTCTAGGTTACATTCTATTTCAGTATGTTTTTGGTGTTATGATTATTAACACTAATGATATGAGT
CCTGCTTTAAGTGCAGGTGACGGTGTTTTATATTATCGTTTGACTGATCGCTATCATATTAATGATGTGGTGGTC
TATGAGGTTGATAACACTTTGAAAGTTGGTCGAATTGTCGCTCAAGCTGGCGATGAGGTTAGTTTTACGCAAGAA
GGAGGACTGTTGATTAATGGGCATCCACCAGAAAAAGAGGTCCCTTACCTGACGTATCCTCACTCAAGTGGCCCA
AACTTTCCCTATAAAGTTCCTACGGGTAAGTATTTCATATTGAATGATTATCGTGAAGAACGTTTGGACAGTCGT
TATTATGGGGCGTTACCCGTCAATCAAATAAAAGGGAAAATCTCAACTCTATTAAGAGTGAGAGGAATT

SEQ ID NO: 93

MTNYLNRLNENSLFKAFIRLVLKISIIGFLGYILFQYVFGVMIINTNDMSPALSAGDGVLYYRLTDRYHINDVVV
YEVDNTLKVGRIVAQAGDEVSFTQEGGLLINGHPPEKEVPYLTYPHSSGPNFPYKVPTGKYFILNDYREERLDSR
YYGALPVNQIKGKISTLLRVRGI

Orf 80 is thought to to be a fimbrial protein. An example of the nucleotide sequence encoding the fimbrial protein (SEQ ID NO: 94) and a fimbrial protein amino acid sequence (SEQ ID NO: 95) are set forth below.

SEQ ID NO: 94

TTGGAGAGAGAAAAAATGAAAAAAAACAAATTATTACTTGCTACTGCAATCTTAGCAACTGCTTTAGGAACAGCT
TCTTTAAATCAAAACGTAAAAGCTGAGACGGCAGGGGTTGTAACAGGAAAATCACTACAAGTTACAAAGACAATG
ACTTATGATGATGAAGAGGTGTTAATGCCCGAAACCGCCTTTACTTTTACTATAGAGCCTGATATGACTGCAAGT
GGAAAAGAAGGCAGCCTAGATATTAAAAATGGAATTGTAGAAGGCTTAGACAAACAAGTAACAGTAAAATATAAG
AATACAGATAAAACATCTCAAAAAACTAAAATAGCACAATTTGATTTTTCTAAGGTTAAATTTCCAGCTATAGGT
GTTTACCGCTATATGGTTTCAGAGAAAAACGATAAAAAAGACGGAATTACGTACGATGATAAAAAGTGGACTGTA
GATGTTTATGTTGGGAATAAGGCCAATAACGAAGAAGGTTTCGAAGTTCTATATATTGTATCAAAAGAAGGTACT
TCTAGTACTAAAAAACCAATTGAATTTACAAACTCTATTAAAACTACTTCCTTAAAAATTGAAAAACAAATAACT
GGCAATGCAGGAGATCGTAAAAAATCATTCAACTTCACATTAACATTACAACCAAGTGAATATTATAAAACTGGA
TCAGTTGTGAAAATCGAACAGGATGGAAGTAAAAAAGATGTGACGATAGGAACGCCTTACAAATTTACTTTGGGA
CACGGTAAGAGTGTCATGTTATCGAAATTACCAATTGGTATCAATTACTATCTTAGTGAAGACGAAGCGAATAAA
GACGGCTACACTACAACGGCAACATTAAAAGAACAAGGCAAAGAAAAGAGTTCCGATTTCACTTTGAGTACTCAA
AACCAGAAAACAGACGAATCTGCTGACGAAATCGTTGTCACAAATAAGCGTGACACTCAAGTTCCAACTGGTGTT
GTAGGGACCCTTGCTCCATTTGCAGTTCTTAGCATTGTGGCTATTGGTGGAGTTATCTATATTACAAAACGTAAA
AAAGCT

SEQ ID NO: 95

MEREKMKKNKLLLATAILATALGTASLNQNVKAETAGVVTGKSLQVTKTMTYDDEEVLMPETAFTFTIEPDMTAS
GKEGSLDIKNGIVEGLDKQVTVKYKNTDKTSQKTKIAQFDFSKVKFPAIGVYRYMVSEKNDKKDGITYDDKKWTV
DVYVGNKANNEEGFEVLYIVSKEGTSSTKKPIEFTNSIKTTSLKIEKQITGNAGDRKKSFNFTLTLQPSEYYKTG
SVVKIEQDGSKKDVTIGTPYKFTLGHGKSVMLSKLPIGINYYLSEDEANKDGYTTTATLKEQGKEKSSDFTLSTQ
NQKTDESADEIVVTNKRDT*QVPTG*VVGTLAPFAVLSIVAIGGVIYITKRKKA

Orf 82 contains an amino acid motif indicative of a cell wall anchor: **SEQ ID NO: 140** QVPTG (shown in italics in SEQ ID NO: 95, above). In some recombinant host cell systems, it may be preferable to remove this motif to facilitate secretion of a recombinant Orf 82 protein from the host cell. Alternatively, in other recombinant host cell systems, it may be preferable to use the cell wall anchor motif to anchor the recombinantly expressed protein to the cell wall. The extracellular domain of the expressed protein may be cleaved during purification or the recombinant protein may be left attached to either inactivated host cells or cell membranes in the final composition.

An E box containing a conserved glutamic residue has been identified in Orf 80. The E-box motif is underlined in SEQ ID NO: 95, below. The conserved glutamic acid (E), at amino acid residue 270, is marked in bold. The E box motif, in particular the conserved glutamic acid residue, is thought to be important for the formation of oligomeric pilus-like structures of Orf 80. Preferred fragments of Orf 80 include at least one conserved glutamic acid residue. Preferably, fragments include at least one E box motif.

SEQ ID NO: 95

MEREKMKKNKLLLATAILATALGTASLNQNVKAETAGVVTGKSLQVTKTMTYDDEEVLMPETAFTFTIEPDMTAS
GKEGSLDIKNGIVEGLDKQVTVKYKNTDKTSQKTKIAQFDFSKVKFPAIGVYRYMVSEKNDKKDGITYDDKKWTV
DVYVGNKANNEEGFEVLYIVSKEGTSSTKKPIEFTNSIKTTSLKIEKQITGNAGDRKKSFNFTLTLQPSEYYKTG
SVVKIEQDGSKKDVTIGTPYKFTLGHGKSVMLSKLPIGINYYLS**E**DEANKDGYTTTATLKEQGKEKSSDFTLSTQ
NQKTDESADEIVVTNKRDTQVPTGVVGTLAPFAVLSIVAIGGVIYITKRKKA

Orf 81 is thought to to be a SrtC2 type sortase. An example of the nucleotide sequence encoding the SrtC2 sortase (SEQ ID NO: 96) and a SrtC2 sortase amino acid sequence (SEQ ID NO: 97) are set forth below.

SEQ ID NO: 96

GTGATTAGTCAAAGAATGATGATGACAATTGTACAGGTTATCAATAAAGCCATTGATACTCTCATTCTTATCTTT
TGTTTAGTCGTACTATTTTTAGCTGGTTTTGGTTTGTGGGATTCTTATCATCTCTATCAACAAGCAGACGCTTCT
AATTTCAAAAAATTTAAAACAGCTCAACAACAGCCTAAATTTGAAGACTTGTTAGCTTTGAATGAGGATGTCATT
GGTTGGTTAAATATCCCAGGGACTCATATTGATTATCCTCTAGTTCAGGGAAAAACGAATTTAGAGTATATTAAT
AAAGCAGTTGATGGCAGTGTTGCCATGTCTGGTAGTTTATTTTTAGATACACGGAATCATAATGATTTTACGGAC
GATTACTCTCTGATTTATGGCCATCATATGGCAGGTAATGCCATGTTTGGCGAAATTCCAAAATTTTTAAAAAAG
GATTTTTTCAACAAACATAATAAAGCTATCATTGAAACAAAAGAGAGAAAAAAACTAACCGTCACTATTTTTGCT
TGTCTCAAGACAGATGCCTTTGACCAGTTAGTTTTTAATCCTAATGCTATTACCAATCAAGACCAACAAAAGCAG
CTCGTTGATTATATCAGTAAAAGATCAAAACAATTTAAACCTGTTAAATTGAAGCATCATACAAAGTTCGTTGCT
TTTTCAACGTGTGAAAATTTTTCTACTGACAATCGTGTTATCGTTGTCGGTACTATTCAAGAA

SEQ ID NO: 97

MISQRMMMTIVQVINKAIDTLILIFCLVVLFLAGFGLWDSYHLYQQADASNFKKFKTAQQQPKFEDLLALNEDVI
GWLNIPGTHIDYPLVQGKTNLEYINKAVDGSVAMSGSLFLDTRNHNDFTDDYSLIYGHHMAGNAMFGEIPKFLKK
DFFNKHNKAIIETKERKKLTVTIFACLKTDAFDQLVFNPNAITNQDQQKQLVDYISKRSKQFKPVKLKHHTKFVA
FSTCENFSTDNRVIVVGTIQE

Orf 82 is referred to as a hypothetical protein. It contains a sortase substrate motif LPXAG shown in italics in SEQ ID NO: 99. An example of the nucleotide sequence encoding the hypothetical protein (SEQ ID NO: 98) and a hypothetical protein amino acid sequence (SEQ ID NO: 99) are set forth below.

SEQ ID NO: 98

TTGCTTTTTCAACGTGTGAAAATTTTTCTACTGACAATCGTGTTATCGTTGTCGGTACTATTCAAGAATAACGAA
AGGAGGAGACTTTTGAGAAAATATTGGAAAATGTTATTTTCTGTCGTAATGATATTAACCATGCTGGCCTTTAAT
CAGACTGTTTTAGCAAAAGACAGCACTGTTCAAACTAGCATTAGTGTCGAAAATGTCTTAGAGAGAGCAGGCGAT
AGTACCCCATTTTCGGTTGCATTAGAATCAATTGATGCGATGAAAACAATAGACGAAATAACAATTGCTGGTTCT
GGAAAAGCAAGCTTTTCCCCTCTGACCTTCACAACAGTTGGGCAATATACTTATCGTGTTTATCAGAAGCCTTCA
CAAAATAAAGATTATCAAGCAGATACTACTGTATTTGACGTTCTTGTCTATGTGACCTATGATGAAGATGGGACT
CTAGTCGCAAAAGTTATTTCTCGAAGGGCTGGAGACGAAGAAAAATCAGCGATTACTTTTAAGCCCAAACGGTTA
GTAAAACCAATACCGCCTAGACAACCTAACATCCCTAAAACCCCATTACCATTAGCTGGTGAAGTAAAAAGTTTA
TTGGGTATCTTAAGTATCGTATTACTGGGGTTACTAGTTCTTCTTTATGTTAAAAAACTGAAGAGTAGGCTA

SEQ ID NO: 99

MLFQRVKIFLLTIVLSLSVLFKNNERRRLLRKYWKMLFSVVMILTMLAFNQTVLAKDSTVQTSISVENVLERAGD
STPFSVALESIDAMKTIDEITIAGSGKASFSPLTFTTVGQYTYRVYQKPSQNKDYQADTTVFDVLVYVTYDEDGT
LVAKVISRRAGDEEKSAITFKPKRLVKPIPPRQPNIPKTP*LPLAG*EVKSLLGILSIVLLGLLVLLYVKKLKSRL

Orf 82 contains an amino acid motif indicative of a cell wall anchor: **SEQ ID NO: 185** LPLAG (shown in italics in SEQ ID NO: 99, above). In some recombinant host cell systems, it may be preferable to remove this motif to facilitate secretion of a recombinant Orf 82 protein from the host cell. Alternatively, in other recombinant host cell systems, it may be preferable to use the cell wall anchor motif to anchor the recombinantly expressed protein to the cell wall. The extracellular domain of the expressed protein may be cleaved during purification or the recombinant protein may be left attached to either inactivated host cells or cell membranes in the final composition.

A pilin motif, discussed above, containing a conserved lysine (K) residue has also been identified in Orf 82. The pilin motif sequence is underlined in SEQ ID NO: 99, below. Conserved lysine (K) residues are also marked in bold, at amino acid residues 173 and 188. The pilin sequence, in particular the conserved lysine residues, are thought to be important for the formation of oligomeric, pilus-like structures. Preferred fragments of Orf 82 include at least one conserved lysine residue. Preferably, fragments include the pilin sequence.

SEQ ID NO: 99

MLFQRVKIFLLTIVLSLSVLFKNNERRRLLRKYWKMLFSVVMILTMLAFNQTVLAKDSTVQTSISVENVLERAGD
STPFSVALESIDAMKTIDEITIAGSGKASFSPLTFTTVGQYTYRVYQKPSQNKDYQADTTVFDVLVYVTYDEDGT
LVAKVISRRAGDEEKSAITFKP**K**RLVKPIPPRQPNIPKTPLPLAGEVKSLLGILSIVLLGLLVLLYVKKLKSRL

An E box containing a conserved glutamic residue has been identified in Orf 82. The E-box motif is underlined in SEQ ID NO: 99, below. The conserved glutamic acid (E), at amino acid residue 163, is marked in bold. The E box motif, in particular the conserved glutamic acid residue, is thought to be important for the formation of oligomeric pilus-like structures of Orf 82. Preferred fragments of Orf 82 include the conserved glutamic acid residue. Preferably, fragments include the E box motif.

SEQ ID NO: 99

MLFQRVKIFLLTIVLSLSVLFKNNERRRLLRKYWKMLFSVVMILTMLAFNQTVLAKDSTVQTSISVENVLERAGD
STPFSVALESIDAMKTIDEITIAGSGKASFSPLTFTTVGQYTYRVYQKPSQNKDYQADTTVFDVLVYVTYDEDGT
LVAKVISRRAGD**E**EKSAITFKPKRLVKPIPPRQPNIPKTPLPLAGEVKSLLGILSIVLLGLLVLLYVKKLKSRL

Orf 83 is thought to to be a multiple sugar metabolism regulator protein. An example of a nucleotide sequence encoding the sugar metabolism regulator protein (SEQ ID NO: 100) and a sugar metabolism regulator protein amino acid sequence (SEQ ID NO: 101) are set forth below.

SEQ ID NO: 100

ATGATACAACTAAGGATGGGGGCAATCTATCAAATGGTTATATTCGATTTAAAACATGTGCAAACATTACACAGC
TTGTCTCAATTACCTATTTCAGTGATGTCACAAGATAAGGCACTTATTCAAGTATATGGTAATGACGACTATTTA
TTATGTTACTATCAATTTTTAAAGCATCTAGCTATTCCTCAAGCTGCACAAGATGTTATTTTTTATGAGGGTTTA
TTTGAAGAGTCCTTTATGATTTTTCCTCTTTGTCACTACATTATTGCCATTGGACCTTTCTATCCTTATTCACTT
AATAAAGACTATCAGGAACAATTAGCTAATAATTTTTTAAAACATTCTTCTCATCGTAGCAAAGAAGAGCTCTTG
TCCTATATGGCACTTGTCCCACATTTTCCAATTAATAATGTGCGGAACCTTTTGATAGCTATTGACGCTTTTTTT
GACACACAATTTGAGACGACTTGCCAACAAACGATTCATCAATTGTTGCAGCATTCAAAACAGATGACTGCTGAT
CCTGATATCATTCATCGCCTTAAGCATATTAGCAAAGCATCTAGCCAATTACCGCCTGTTTTAGAGCACCTAAAT
CATATTATGGATCTGGTAAAGCTAGGCAATCCACAATTGCTCAAGCAAGAAATCAATCGCATCCCCTTATCAAGT
ATCACCTCATCTTCTATTTCTGCTCTAAGGGCGGAAAAGAACCTCACTGTTATCTATTTAACTAGGTTACTGGAA
TTCAGTTTTGTAGAAAATACTGACGTAGCAAAGCATTATAGCCTTGTCAAATACTACATGGCCTTAAATGAAGAA
GCGAGTGACTTGCTCAAAGTTTTGAGAATTCGCTGTGCAGCTATCATCCATTTTTCCGAATCATTAACCAATAAA
AGTATTTCTGATAAACGTCAAATGTACAATAGTGTGCTTCATTATGTCGATAGTCACCTGTATTCCAAATTAAAG

GTATCTGATATCGCTAAGCGCCTATATGTTTCCGAATCTCACTTACGTTCAGTCTTTAAAAAATACTCAAATGTT
TCCTTACAACATTATATTCTAAGTACAAAAATCAAAGAAGCTCAACTACTCTTAAAACGAGGAATTCCTGTTGGA
GAAGTGGCTAAAAGCTTATATTTTTATGACACTACCCATTTTCATAAAATCTTTAAAAAATACACGGGTATTTCT
TCAAAAGACTATCTTGCTAAATACCGAGATAATATT

SEQ ID NO: 101

MIQLRMGAIYQMVIFDLKHVQTLHSLSQLPISVMSQDKALIQVYGNDDYLLCYYQFLKHLAIPQAAQDVIFYEGL
FEESFMIFPLCHYIIAIGPFYPYSLNKDYQEQLANNFLKHSSHRSKEELLSYMALVPHFPINNVRNLLIAIDAFF
DTQFETTCQQTIHQLLQHSKQMTADPDIIHRLKHISKASSQLPPVLEHLNHIMDLVKLGNPQLLKQEINRIPLSS
ITSSSISALRAEKNLTVIYLTRLLEFSFVENTDVAKHYSLVKYYMALNEEASDLLKVLRIRCAAIIHFSESLTNK
SISDKRQMYNSVLHYVDSHLYSKLKVSDIAKRLYVSESHLRSVFKKYSNVSLQHYILSTKIKEAQLLLKRGIPVG
EVAKSLYFYDTTHFHKIFKKYTGISSKDYLAKYRDNI

Orf 84 is thought to to be a F2-like fibronectin-binding protein. An example of a nucleotide sequence encoding the F2-like fibronectin-binding protein (SEQ ID NO: 102) and a F2-like fibronectin-binding protein amino acid sequence (SEQ ID NO: 103) are set forth below.

SEQ ID NO: 102

ATGACACAAAAAAATAGCTATAAGTTAAGCTTCCTGTTATCCCTAACAGGATTTATTTTAGGTTTATTATTGGTT
TTTATAGGATTGTCCGGAGTATCAGTAGGACATGCGGAAACAAGAAATGGAGCAAACAAACAAGGAGCTTTTGAA
ATCAAGAAAAATAAAAGTCAAGAAGAATATAATTATGAAGTTTATGATAACAGAAACATACTTCAGGATGGGGAA
CATAAACTTGAAATAAAAAGAGTTGATGGGACAGGTAAAACTTATCAAGGTTTTTGCTTTCAGTTAACGAAAAAT
TTTCCCACTGCTCAAGGTGTAAGTAAAAAGCTGTATAAAAAATTGAGTAGTAGTGATGAAGAAACACTAAAGCAA
TATGCCTCTAAGTATACAAGTAATAGGAGAGGAGATACTAGTGGTAATCTTAAAAAGCAAATTGCTAAGGTTCTG
ACAGAAGGTTACCCAACTAACAAAAGTGATTGGTTAAATGGATTGACTGAAAACGAAAAAATAGAAGTAACCCAG
GATGCAATTTGGTATTTTACAGAAACGACAGTTCCGGCTGATAGAAGTTATACGAATCGCAACGTAAATAGTCAA
AAAATGAAAGAAGTGTATCAAAAGCTAATTGATACAACAGATATAGATAAATATGAAGATGTACAATTTGATTTA
TTTGTGCCACAAGATACAAACTTACAGGCAGTAATTAGTGTAGAGCCTGTTATCGAAAGCCTTCCTTGGACATCG
TTGAAGCCAATAGCCCAGAAGGATATCACTGCCAAAAAAATCTGGGTAGATGCACCTAAAGAAAAACCAATTATT
TATTTTAAGCTATATAGACAGCTGCCTGGAGAAAAGGAAGTAGCAGTGGATGACGCTGAGCTAAAACAGATAAAT
AGTGAAGGTCAACAAGAAATATCAGTAACTTGGACAAATCAACTTGTTACAGATGAAAAAGGAATGGCTTACATT
TATTCTGTAAAAGAAGTAGATAAAAATGGCGAGTTACTTGAGCCAAAAGATTATATCAAGAAGGAAGATGGACTT
ACAGTTACTAATACTTATGTAAAGCCAACTAGTGGGCACTATGATATAGAAGTGACATTTGGAAATGGACATATT
GATATTACAGAAGATACTACACCAGATATTGTTTCAGGTGAAAACCAAATGAAGCAAATAGAGGGAGAAGATAGT
AAGCCTATTGATGAAGTAACGGAAAATAATTTAATTGAATTTGGTAAAAACACGATGCCAGGTGAAGAAGATGGC
ACAAATTCTAATAAGTATGAAGAAGTCGAAGACTCACGCCCAGTTGATACCTTGTCAGGTTTATCAAGTGAGCAA
GGTCAGTCCGGTGATATGACAATTGAAGAAGATAGTGCTACCCATATTAAATTCTCAAAACGTGATATTGACGGC
AAAGAGTTAGCTGGTGCAACTATGGAGTTGCGTGATTCATCTGGTAAAACTATTAGTACATGGATTTCAGATGGA
CAAGTGAAAGATTTCTACCTGATGCCAGGAAAATATACATTTGTCGAAACCGCAGCACCAGACGGTTATGAGATA
GCAACTGCTATTACCTTTACAGTTAATGAGCAAGGTCAGGTTACTGTAAATGGCAAAGCAACTAAAGGTGACGCT
CATATTGTCATGGTTGATGCTTACAAGCCAACTAAGGGTTCAGGTCAGGTTATTGATATTGAAGAAAAGCTTCCA
GACGAGCAGGGCCATTCTGGCTCAACTACTGAAATAGAAGATAGCAAGTCTTCAGACGTTATCATTGGTGGTCAG
GGGCAGATTGTCGAGACAACAGAGGATACCCAAACTGGCATGCACGGGGATTCTGGTTGTAAAACGGAAGTCGAA
GATACTAAACTAGTACAATCCTTCCACTTTGATAACAAGGAATCAGAAAGTAACTCTGAGATTCCTAAAAAAGAT
AAGCCAAAGAGTAATACTAGTTTACCAGCAACTGGTGAGAAGCAACATAATATGTTCTTTTGGATGGTTACTTCT
TGCTCACTTATTAGTAGTGTTTTTGTAATATCACTAAAACTAAAAAACGCCTATCATCATGT

SEQ ID NO: 103

MTQKNSYKLSFLLSLTGFILGLLLVFIGLSGVSVGHAETRNGANKQGAFEIKKNKSQEEYNYEVYDNRNILQDGE
HKLEIKRVDGTGKTYQGFCFQLTKNFPTAQGVSKKLYKKLSSSDEETLKQYASKYTSNRRGDTSGNLKKQIAKVL
TEGYPTNKSDWLNGLTENEKIEVTQDAIWYFTETTVPADRSYTNRNVNSQKMKEVYQKLIDTTDIDKYEDVQFDL
FVPQDTNLQAVISVEPVIESLPWTSLKPIAQKDITAKKIWVDAPKEKPIIYFKLYRQLPGEKEVAVDDAELKQIN
SEGQQEISVTWTNQLVTDEKGMAYIYSVKEVDKNGELLEPKDYIKKEDGLTVTNTYVKPTSGHYDIEVTFGNGHI
DITEDTTPDIVSGENQMKQIEGEDSKPIDEVTENNLIEFGKNTMPGEEDGTNSNKYEEVEDSRPVDTLSGLSSEQ
GQSGDMTIEEDSATHIKFSKRDIDGKELAGATMELRDSSGKTISTWISDGQVKDFYLMPGKYTFVETAAPDGYEI
ATAITFTVNEQGQVTVNGKATKGDAHIVMVDAYKPTKGSGQVIDIEEKLPDEQGHSGSTTEIEDSKSSDVIIGGQ
GQIVETTEDTQTGMHGDSGCKTEVEDTKLVQSFHFDNKESESNSEIPKKDKPKSNTS*LPATG*EKQHNMFFWMVTS
CSLISSVFVISLKTKKRLSSC

Orf 84 contains an amino acid motif indicative of a cell wall anchor: **SEQ ID NO: 181** LPATG (shown in italics in SEQ ID NO: 103, above). In some recombinant host cell systems, it may be preferable to remove this motif to facilitate secretion of a recombinant Orf 84 protein from the host cell. Alternatively, in other recombinant host cell systems, it may be preferable to use the cell wall anchor motif to anchor the recombinantly expressed protein to the cell wall. The extracellular domain of the expressed protein may be cleaved during purification or the recombinant protein may be left attached to either inactivated host cells or cell membranes in the final composition.

A pilin motif, discussed above, containing a conserved lysine (K) residue has also been identified in Orf 84. The pilin motif sequence is underlined in SEQ ID NO: 103, below. A conserved lysine (K) residue is also marked in bold, at amino acid residue 270. The pilin sequence, in particular the conserved lysine residue, is thought to be important for the formation of oligomeric, pilus-like structures. Preferred fragments of Orf 84 include the conserved lysine residue. Preferably, fragments include the pilin sequence.

SEQ ID NO: 103

MTQKNSYKLSFLLSLTGFILGLLLVFIGLSGVSVGHAETRNGANKQGAFEIKKNKSQEEYNYEVYDNRNILQDGE
HKLEIKRVDGTGKTYQGFCFQLTKNFPTAQGVSKKLYKKLSSSDEETLKQYASKYTSNRRGDTSGNLKKQIAKVL
TEGYPTNKSDWLNGLTENEKIEVTQDAIWYFTETTVPADRSYTNRNVNSQKMKEVYQKLIDTTDIDKYEDVQFDL
FVPQDTNLQAVISVEPVIESLPWTSLKPIAQKDITAKKIWVDAP**K**EKPIIYFKLYRQLPGEKEVAVDDAELKQIN
SEGQQEISVTWTNQLVTDEKGMAYIYSVKEVDKNGELLEPKDYIKKEDGLTVTNTYVKPTSGHYDIEVTFGNGHI
DITEDTTPDIVSGENQMKQIEGEDSKPIDEVTENNLIEFGKNTMPGEEDGTNSNKYEEVEDSRPVDTLSGLSSEQ
GQSGDMTIEEDSATHIKFSKRDIDGKELAGATMELRDSSGKTISTWISDGQVKDFYLMPGKYTFVETAAPDGYEI
ATAITFTVNEQGQVTVNGKATKGDAHIVMVDAYKPTKGSGQVIDIEEKLPDEQGHSGSTTEIEDSKSSDVIIGGQ
GQIVETTEDTQTGMHGDSGCKTEVEDTKLVQSFHFDNKESESNSEIPKKDKPKSNTSLPATGEKQHNMFFWMVTS
CSLISSVFVISLKTKKRLSSC

An E box containing a conserved glutamic residue has been identified in Orf 84. The E-box motif is underlined in SEQ ID NO: 103, below. The conserved glutamic acid (E), at amino acid residue 516, is marked in bold. The E box motif, in particular the conserved glutamic acid residue, is thought to be important for the formation of oligomeric pilus-like structures of Orf 84. Preferred fragments of Orf 84 include the conserved glutamic acid residue. Preferably, fragments include the E box motif.

SEQ ID NO: 103

MTQKNSYKLSFLLSLTGFILGLLLVFIGLSGVSVGHAETRNGANKQGAFEIKKNKSQEEYNYEVYDNRNILQDGE
HKLEIKRVDGTGKTYQGFCFQLTKNFPTAQGVSKKLYKKLSSSDEETLKQYASKYTSNRRGDTSGNLKKQIAKVL
TEGYPTNKSDWLNGLTENEKIEVTQDAIWYFTETTVPADRSYTNRNVNSQKMKEVYQKLIDTTDIDKYEDVQFDL
FVPQDTNLQAVISVEPVIESLPWTSLKPIAQKDITAKKIWVDAPKEKPIIYFKLYRQLPGEKEVAVDDAELKQIN
SEGQQEISVTWTNQLVTDEKGMAYIYSVKEVDKNGELLEPKDYIKKEDGLTVTNTYVKPTSGHYDIEVTFGNGHI
DITEDTTPDIVSGENQMKQIEGEDSKPIDEVTENNLIEFGKNTMPGEEDGTNSNKYEEVEDSRPVDTLSGLSSEQ
GQSGDMTIEEDSATHIKFSKRDIDGKELAGATMELRDSSGKTISTWISDGQVKDFYLMPGKYTFV**E**TAAPDGYEI
ATAITFTVNEQGQVTVNGKATKGDAHIVMVDAYKPTKGSGQVIDIEEKLPDEQGHSGSTTEIEDSKSSDVIIGGQ
GQIVETTEDTQTGMHGDSGCKTEVEDTKLVQSFHFDNKESESNSEIPKKDKPKSNTSLPATGEKQHNMFFWMVTS
CSLISSVFVISLKTKKRLSSC

Examples of GAS AI-3 sequences from M18 strain isolate MGAS8232 are set forth below.

SpyM18_0125 is a negative transcriptional regulator (Nra). An example of SpyM18_0125 is set forth in SEQ ID NO: 72.

SEQ ID NO: 72

MPYVKKKKDSFLVETYLEQSIRDKSELVELLLFKSPTIIFSHVAKQTGLTAVQLKYYCKELDDFFGNNLDITIKKG
KIICCFVKPVKEFYLHQLYDTSTILKLLVFFIKNGTTSQPLIKFSKKYFLSSSSAYRLRESLIKLLREFGLRVSK
NTIVGEEYRIRYLIAMLYSKFGIVIYPLDHLDNQIIYRFLSQSATNLRTSPWLEEPFSFYNMLLALS

SpyM18_0126 is thought to be a collagen binding protein (CBP). An example of SpyM18_0126 is set forth in SEQ ID NO: 73.

SEQ ID NO: 73
MQKRDKTNYGSANNKRRQTTIGLLKVFLTFVALIGIVGFSIRAFGAEEQSTETKKTSVIIRKYAEGDYSKLLEGA
TLKLAQIEGSGFQEQSFESSTSGQKLQLSDGTYILTETKSPQGYEIAEPITFKVTAGKVFIKGKDGQFVENQNKE
VAEPYSVTAYNDFDDSGFINPKTFTPYGKFYYAKNANGTSQVVYCFNVDLHSPPDSLDKGETIDPDFNEGKEIKY
THILGADLFSYANNPRASTNDELLSQVKKVLEKGYRDDSTTYANLTSVEFRAATQLAIYYFTDSVDLDNLADYHG
FGALTTEALNATKEIVAYAEDRANLPNISNLDFYVPNSNKYQSLIGTQYHPESLVDIIRMEDKQAPIIPITHKLT
ISKTVTGTIADKKKEFNFEIHLKSSDGQAISGTYPTNSGELTVTDGKATFTLKDGESLIVEGLPSGYSYEITETG
ASDYEVSVNGKNAPDGKATKASVKEDETITFENRKDL*VPPTG*LTTDGAIYLWLLLLVLLGLWVWLIGRKGLKND

SpyM18_0126 contains an amino acid motif indicative of a cell wall anchor: **SEQ ID NO: 184** VPPTG (shown in italics in SEQ ID NO: 73, above). In some recombinant host cell systems, it may be preferable to remove this motif to facilitate secretion of a recombinant SpyM18_0126 protein from the host cell. Alternatively, in other recombinant host cell systems, it may be preferable to use the cell wall anchor motif to anchor the recombinantly expressed protein to the cell wall. The extracellular domain of the expressed protein may be cleaved during purification or the recombinant protein may be left attached to either inactivated host cells or cell membranes in the final composition.

A pilin motif, discussed above, containing a conserved lysine (K) residue has also been identified in SpyM18_0126. The pilin motif sequence is underlined in SEQ ID NO: 73, below. Conserved lysine (K) residues are also marked in bold, at amino acid residues 172 and 179. The pilin sequence, in particular the conserved lysine residues, are thought to be important for the formation of oligomeric, pilus-like structures. Preferred fragments of SpyM18_0126 include at least one conserved lysine residue. Preferably, fragments include the pilin sequence.

SEQ ID NO: 73
MQKRDKTNYGSANNKRRQTTIGLLKVFLTFVALIGIVGFSIRAFGAEEQSTETKKTSVIIRKYAEGDYSKLLEGA
TLKLAQIEGSGFQEQSFESSTSGQKLQLSDGTYILTETKSPQGYEIAEPITFKVTAGKVFIKGKDGQFVENQNKE
VAEPYSVTAYNDFDDSGFINP**K**TFTPYG**K**FYYAKNANGTSQVVYCFNVDLHSPPDSLDKGETIDPDFNEGKEIKY
THILGADLFSYANNPRASTNDELLSQVKKVLEKGYRDDSTTYANLTSVEFRAATQLAIYYFTDSVDLDNLADYHG
FGALTTEALNATKEIVAYAEDRANLPNISNLDFYVPNSNKYQSLIGTQYHPESLVDIIRMEDKQAPIIPITHKLT
ISKTVTGTIADKKKEFNFEIHLKSSDGQAISGTYPTNSGELTVTDGKATFTLKDGESLIVEGLPSGYSYEITETG
ASDYEVSVNGKNAPDGKATKASVKEDETITFENRKDLVPPTGLTTDGAIYLWLLLLVLLGLWVWLIGRKGLKND

Three E boxes containing conserved glutamic residues have been identified in SpyM18_0126. The E-box motifs are underlined in SEQ ID NO: 73, below. The conserved glutamic acid (E) residues, at amino acid residues 112, 257, and 415, are marked in bold. The E box motifs, in particular the conserved glutamic acid residues, are thought to be important for the formation of oligomeric pilus-like structures of SpyM18_0126. Preferred fragments of SpyM18_0126 include at least one conserved glutamic acid residue. Preferably, fragments include at least one E box motif.

SEQ ID NO: 73
MQKRDKTNYGSANNKRRQTTIGLLKVFLTFVALIGIVGFSIRAFGAEEQSTETKKTSVIIRKYAEGDYSKLLEGA
TLKLAQIEGSGFQEQSFESSTSGQKLQLSDGTYILT**E**TKSPQGYEIAEPITFKVTAGKVFIKGKDGQFVENQNKE

VAEPYSVTAYNDFDDSGFINPKTEITPYGKFYYAKNANGTSQVVYCFNVDLHSPPDSLDKGETIDPDFNEGKEIKY
THILGADLFSYANNPRASTNDELLSQVKKVL**E**KGYRDDSTTYANLTSVEFRAATQLAIYYFTDSVDLDNLADYHG
FGALTTEALNATKEIVAYAEDRANLPNISNLDFYVPNSNKYQSLIGTQYHPESLVDIIRMEDKQAPIIPITHKLT
ISKTVTGTIADKKKEFNFEIHLKSSDGQAISGTYPTNSG**E**LTVTDGKATFTLKDGESLIVEGLPSGYSYEITETG
ASDYEVSVNGKNAPDGKATKASVKEDETITFENRKDLVPPTGLTTDGAIYLWLLLLVLLGLWVWLIGRKGLKND

SpyM18_0127 is a LepA protein. An example of SpyM18_0127 is shown in SEQ ID NO: 74.

SEQ ID NO: 74
MTNYLNRLNENPLFKAFIRLVLKISIIGFLGYILFQYIFGVMIINTNVMSPALSAGDGILYYRLTDRYHINDVVV
YEVDNTLKVGRIVAQAGDEVSFTQEGGLLINGHPPEKEVPYLTYPHSSGPNFPYKVPTGTYFILNDYREERLDSR
YYGALPINQIKGKISTLLRVRGI

SpyM18_0128 is thought to be a fimbrial protein. An example of SypM18_0128 is shown in SEQ ID NO: 75.

SEQ ID NO: 75
MKKNKLLLATAILATALGTASLNQNVKAETAGVIDGSTLVVKKTFPSYTDDKVLMPKADYTFKVEADDNAKGKTK
DGLDIKPGVIDGLENTKTIHYGNSDKTTAKEKSVNFDFANVKFPGVGVYRYTVSEVNGNKAGIAYDSQQWTVDVY
VVNREDGGFEAKYIVSTEGGQSDKKPVLFKNFFDTTSLKVTKKVTGNTGEHQRSFSFTLLLTPNECFEKGQVVNI
LQGGETKKVVIGEEYSFTLKDKESVTLSQLPVGIEYKVTEEDVTKDGYKTSATLKDGDVTDGYNLGDSKTTDKST
DEIVVTNKRDT*QVPTG*VVGTLAPFAVLSIVAIGGVIYITKRKKA

SpyM18_0128 contains an amino acid motif indicative of a cell wall anchor: **SEQ ID NO: 140** QVPTG (shown in italics in SEQ ID NO: 75, above). In some recombinant host cell systems, it may be preferable to remove this motif to facilitate secretion of a recombinant SpyM18_0128 protein from the host cell. Alternatively, in other recombinant host cell systems, it may be preferable to use the cell wall anchor motif to anchor the recombinantly expressed protein to the cell wall. The extracellular domain of the expressed protein may be cleaved during purification or the recombinant protein may be left attached to either inactivated host cells or cell membranes in the final composition.

A pilin motif, discussed above, containing a conserved lysine (K) residue has also been identified in SpyM18_0128. The pilin motif sequence is underlined in SEQ ID NO: 75, below. A conserved lysine (K) residue is also marked in bold, at amino acid residue 57. The pilin sequence, in particular the conserved lysine residue, is thought to be important for the formation of oligomeric, pilus-like structures. Preferred fragments of SpyM18_0128 include the conserved lysine residue. Preferably, fragments include at least one pilin sequence.

SEQ ID NO: 75
MKKNKLLLATAILATALGTASLNQNVKAETAGVIDGSTLVVKKTFPSYTDDKVLMP**K**ADYTFKVEADDNAKGKTK
DGLDIKPGVIDGLENTKTIHYGNSDKTTAKEKSVNFDFANVKFPGVGVYRYTVSEVNGNKAGIAYDSQQWTVDVY
VVNREDGGFEAKYIVSTEGGQSDKKPVLFKNFFDTTSLKVTKKVTGNTGEHQRSFSFTLLLTPNECFEKGQVVNI
LQGGETKKVVIGEEYSFTLKDKESVTLSQLPVGIEYKVTEEDVTKDGYKTSATLKDGDVTDGYNLGDSKTTDKST
DEIVVTNKRDTQVPTGVVGTLAPFAVLSIVAIGGVIYITKRKKA

An E box containing a conserved glutamic residue has been identified in SpyM18_0128. The E-box motif is underlined in SEQ ID NO: 75, below. The conserved glutamic acid (E), at amino acid residue 266, is marked in bold. The E box motif, in particular the conserved glutamic acid residue, is thought to be important for the formation of oligomeric pilus-like structures of SpyM18_0128.

Preferred fragments of SpyM18_0128 include the conserved glutamic acid residue. Preferably, fragments include the E box motif.

SEQ ID NO: 75
MKKNKLLLATAILATALGTASLNQNVKAETAGVIDGSTLVVKKTFPSYTDDKVLMPKADYTFKVEADDNAKGKTK
DGLDIKPGVIDGLENTKTIHYGNSDKTTAKEKSVNFDFANVKFPGVGVYRYTVSEVNGNKAGIAYDSQQWTVDVY
VVNREDGGFEAKYIVSTEGGQSDKKPVLFKNFFDTTSLKVTKKVTGNTGEHQRSFSFTLLLTPNECFEKGQVVNI
LQGGETKKVVIGEEYSFTLKDKESVTLSQLPVGIEYKVTE**E**DVTKDGYKTSATLKDGDVTDGYNLGDSKTTDKST
DEIVVTNKRDTQVPTGVVGTLAPFAVLSIVAIGGVIYITKRKKA

SpyM18_0129 is a SrtC2 type sortase. An example of SpyM18_0129 is shown in SEQ ID NO: 76

SEQ ID NO: 76
MISQRMMMTIVQVINKAIDTLILIFCLVVLFLAGFGLWDSYHLYQQADASNFKKFKTAQQQPKFEDLLALNEDVI
GWLNIPGTHMDYPLVQGKTNLEYINKAVDGSVAMSGSLFLDTRNHNDFTDDYSLIYGHHMAGNAMFGEIPKFLKK
DFFNKHNKAIIETKERKKLTVTIFACLKTDAFDQLVFNPNAITNQDQQRQLVDYISKRSKQFKPVKLKHHTKFVA
FSTCENFSTDNRVIVVGTIQE

SpyM18_0130 is referred to as a hypothetical protein. An example of SpyM18_0130 is shown in SEQ ID NO: 77.

SEQ ID NO: 77
MRKYWKMLFSVVMILTMLAFNQTVLAKDSTVQTSISVENVLERAGDSTSFSVALESIDAMKTIDEITIAGSGKAS
FSPLTFTTVGQYTYRVYQKPSQNKDYQADTTVFDVLVYVTYDEDGTLVAKVISRRAGDEEKSAITFKPKRLVKPI
PPRQPDIPKTP*LPLAG*EVKSLLGILSIVLLGLLVLLYVKKLKSRL

SpyM18_0130 contains an amino acid motif indicative of a cell wall anchor: **SEQ ID NO: 185** LPLAG (shown in italics in SEQ ID NO: 77, above). In some recombinant host cell systems, it may be preferable to remove this motif to facilitate secretion of a recombinant SpyM18_0130 protein from the host cell. Alternatively, in other recombinant host cell systems, it may be preferable to use the cell wall anchor motif to anchor the recombinantly expressed protein to the cell wall. The extracellular domain of the expressed protein may be cleaved during purification or the recombinant protein may be left attached to either inactivated host cells or cell membranes in the final composition.

A pilin motif, discussed above, containing a conserved lysine (K) residue has also been identified in SpyM18_0130. The pilin motif sequence is underlined in SEQ ID NO: 77, below. Conserved lysine (K) residues are also marked in bold, at amino acid residues 144, 159, and 169. The pilin sequence, in particular the conserved lysine residues, are thought to be important for the formation of oligomeric, pilus-like structures. Preferred fragments of SpyM18_0130 include at least one conserved lysine residue. Preferably, fragments include the pilin sequence.

SEQ ID NO: 77
MRKYWKMLFSVVMILTMLAFNQTVLAKDSTVQTSISVENVLERAGDSTSFSVALESIDAMKTIDEITIAGSGKAS
FSPLTFTTVGQYTYRVYQKPSQNKDYQADTTVFDVLVYVTYDEDGTLVAKVISRRAGDEEKSAITFKP**K**RLVKPI
PPRQPDIP**K**TPLPLAGEV**K**SLLGILSIVLLGLLVLLYVKKLKSRL

An E box containing a conserved glutamic residue has been identified in SpyM18_0130. The E-box motif is underlined in SEQ ID NO: 77, below. The conserved glutamic acid (E), at amino acid residue 134, is marked in bold. The E box motif, in particular the conserved glutamic acid residue, is thought to be important for the formation of oligomeric pilus-like structures of SpyM18_0130. Preferred fragments of SpyM18_0130 include the conserved glutamic acid residue. Preferably, fragments include the E box motif.

SEQ ID NO: 77
MRKYWKMLFSVVMILTMLAFNQTVLAKDSTVQTSISVENVLERAGDSTSFSVALESIDAMKTIDEITIAGSGKAS
FSPLTFTTVGQYTYRVYQKPSQNKDYQADTTVFDVLVYVTYDEDGTLVAKVISRRAGDEEKSAITFKPKRLVKPI
PPRQPDIPKTPLPLAGEVKSLLGILSIVLLGLLVLLYVKKLKSRL

SpyM18_0131 is referred to as a putative multiple sugar metabolism regulator. An example of SpyM18_0131 is set forth in SEQ ID NO: 78.

SEQ ID NO: 78
MAIFDLKHVQTLHSLSQLPISVMSQDKALIQVYGNDDYLLCYYQFLKHLAIPQAAQDVIFYEGLFEESFMIFPLC
HYIIAIGPFYPYSLNKDYQEQLANNCLKHSSHRSKEELLSYMALVPHFPINNVRNLLIAIDAFFDTQFETTCQQT
IHQLLQHSKQMTADPDIIHRLKHISKASSQLPPVLEHLNHIMDLVKLGNPQLLKQEINRIPLSSITSSSISALRA
EKNLTVIYLTRLLEFSFVENTDVAKHYSLVKYYMALNEEASDLLKVLRIRCAAIIHFSESLTNKSISDKRQMYNS
VLHYVDSHLYSKLKVSDIAKRLYVSESHLRSVFKKYSNVSLQHYILSTKIKEAQLLLKRGIPVGEVAKSLYFYDT
THFHKIFKKYTGISSKDYLAKYRDNI

SpyM18_0132 is a F2 like fibronectic-binding protein. An example of SpyM18_0132 is set forth in SEQ ID NO: 79.

SEQ ID NO: 79
MTQKNSYKLSFLLSLTGFILGLLLVFIGLSGVSVGHAETRNGANKQGAFEIKKNKSQEEYNYEVYDNRNILQDGE
HKLEIKRVDGTGKTYQGFCFQLTKNFPTAQGVSKKLYKKLSSSDEETLKQYASKYTSNRRGDTSGNLKKQIAKVL
TEGYPTNKSDWLNGLTENEKIEVTQDAIWYFTETTVPADRSYTNRNVNSQKMKEVYQKLIDTTDIDKYEDVQFDL
FVPQDTNLQAVISVEPVIESLPWTSLKPIAQKDITAKKIWVDAPKEKPIIYFKLYRQLPGEKEVAVDDAELKQIN
SEGQQEISVTWTNQLVTDEKGMAYIYSVKEVDKNGELLEPKDYIKKEDGLTVTNTYVKPTSGHYDIEVTFGNGHI
DITEDTTPDIVSGENQMKQIEGEDSKPIDEVTENNLIEFGKNTMPGEEDGTNSNKYEEVEDSRPVDTLSGLSSEQ
GQSGDMTIEEDSATHIKFSKRDIDGKELAGATMELRDSSGKTISTWISDGQVKDFYLMPGKYTFVETAAPDGYEI
ATAITFTVNEQGQVTVNGKATKGDAHIVMVDAYKPTKGSGQVIDIEEKLPDEQGHSGSTTEIEDSKSSDVIIGGQ
GQIVETTEDTQTGMHGDSGCKTEVEDTKLVQSFHFDNKESESNSEIPKKDKPKSNTS*LPATG*EKQHNMFFWMVTS
CSLISSVFVISLKTKKRLSSC

SpyM18_0132 contains an amino acid motif indicative of a cell wall anchor: **SEQ ID NO: 180** LPATG (shown in italics in SEQ ID NO: 79, above). In some recombinant host cell systems, it may be preferable to remove this motif to facilitate secretion of a recombinant SpyM18_0132 protein from the host cell. Alternatively, in other recombinant host cell systems, it may be preferable to use the cell wall anchor motif to anchor the recombinantly expressed protein to the cell wall. The extracellular domain of the expressed protein may be cleaved during purification or the recombinant protein may be left attached to either inactivated host cells or cell membranes in the final composition.

A pilin motif, discussed above, containing a conserved lysine (K) residue has also been identified in SpyM18_0132. The pilin motif sequence is underlined in SEQ ID NO: 79, below. A conserved lysine (K) residue is also marked in bold, at amino acid residue 270. The pilin sequence, in particular the conserved lysine residue, is thought to be important for the formation of oligomeric, pilus-like structures. Preferred fragments of SpyM18_0132 include the conserved lysine residue. Preferably, fragments include the pilin sequence.

SEQ ID NO: 79

MTQKNSYKLSFLLSLTGFILGLLLVFIGLSGVSVGHAETRNGANKQGAFEIKKNKSQEEYNYEVYDNRNILQDGE
HKLEIKRVDGTGKTYQGFCFQLTKNFPTAQGVSKKLYKKLSSSDEETLKQYASKYTSNRRGDTSGNLKKQIAKVL
TEGYPTNKSDWLNGLTENEKIEVTQDAIWYFTETTVPADRSYTNRNVNSQKMKEVYQKLIDTTDIDKYEDVQFDL
FVPQDTNLQAVISVEPVIESLPWTSLKPIAQKDITAKKIWVDAPKEKPIIYFKLYRQLPGEKEVAVDDAELKQIN
SEGQQEISVTWTNQLVTDEKGMAYIYSVKEVDKNGELLEPKDYIKKEDGLTVTNTYVKPTSGHYDIEVTFGNGHI
DITEDTTPDIVSGENQMKQIEGEDSKPIDEVTENNLIEFGKNTMPGEEDGTNSNKYEEVEDSRPVDTLSGLSSEQ
GQSGDMTIEEDSATHIKFSKRDIDGKELAGATMELRDSSGKTISTWISDGQVKDFYLMPGKYTFVETAAPDGYEI
ATAITFTVNEQGQVTVNGKATKGDAHIVMVDAYKPTKGSGQVIDIEEKLPDEQGHSGSTTEIEDSKSSDVIIGGQ
GQIVETTEDTQTGMHGDSGCKTEVEDTKLVQSFHFDNKESESNSEIPKKDKPKSNTSLPATGEKQHNMFFWMVTS
CSLISSVFVISLKTKKRLSSC

An E box containing a conserved glutamic residue has been identified in SpyM18_0132. The E-box motif is underlined in SEQ ID NO: 79, below. The conserved glutamic acid (E), at amino acid residue 516, is marked in bold. The E box motif, in particular the conserved glutamic acid residue, is thought to be important for the formation of oligomeric pilus-like structures of SpyM18_0132. Preferred fragments of SpyM18_0132 include the conserved glutamic acid residue. Preferably, fragments include the E box motif.

SEQ ID NO: 79

MTQKNSYKLSFLLSLTGFILGLLLVFIGLSGVSVGHAETRNGANKQGAFEIKKNKSQEEYNYEVYDNRNILQDGE
HKLEIKRVDGTGKTYQGFCFQLTKNFPTAQGVSKKLYKKLSSSDEETLKQYASKYTSNRRGDTSGNLKKQIAKVL
TEGYPTNKSDWLNGLTENEKIEVTQDAIWYFTETTVPADRSYTNRNVNSQKMKEVYQKLIDTTDIDKYEDVQFDL
FVPQDTNLQAVISVEPVIESLPWTSLKPIAQKDITAKKIWVDAPKEKPIIYFKLYRQLPGEKEVAVDDAELKQIN
SEGQQEISVTWTNQLVTDEKGMAYIYSVKEVDKNGELLEPKDYIKKEDGLTVTNTYVKPTSGHYDIEVTFGNGHI
DITEDTTPDIVSGENQMKQIEGEDSKPIDEVTENNLIEFGKNTMPGEEDGTNSNKYEEVEDSRPVDTLSGLSSEQ
GQSGDMTIEEDSATHIKFSKRDIDGKELAGATMELRDSSGKTISTWISDGQVKDFYLMPGKYTFV**E**TAAPDGYEI
ATAITFTVNEQGQVTVNGKATKGDAHIVMVDAYKPTKGSGQVIDIEEKLPDEQGHSGSTTEIEDSKSSDVIIGGQ
GQIVETTEDTQTGMHGDSGCKTEVEDTKLVQSFHFDNKESESNSEIPKKDKPKSNTSLPATGEKQHNMFFWMVTS
CSLISSVFVISLKTKKRLSSC

Examples of GAS AI-3 sequences from M49 strain isolate 591 are set forth below.

SpyoM01000156 is a negative transcriptional regulator (Nra). An example of SpyoM01000156 is set forth in SEQ ID NO: 243.

SEQ ID NO: 243

MPYVKKKKDSFLVETYLEQSIRDKSELVLLLFKSPTIIFSHVAKQTGLTAVQLKYYCKELDDFFGNNLDI
TIKKGKIICCFVKPVKEFYLHQLYDTSTILKLLVFFIKNGTSSQPLIKFSKKYFLSSSSAYRLRESLIKL
LREFGLRVSKNTIVGEEYRIRYLIAMLYSKFGIVIYPLDHLDNQIIYRFLSQSATNLRTSPWLEEPFSFY
NMLLALSWKRHQFAVSIPQTRIFRQLKKLFIYDCLTRSSRQVIENAFSLTFSQGDLDYLFLIYITTNNSF
ASLQWTPQHIETCCHIFEKNDTFRLLLEPILKRLPQLNHSKQDLIKALMYFSKSFLFNLQHFVIEIPSFS
LPTYTGNSNLYKALKNIVNQWLAQLPGKRHLNEKHLQLFCSHIEQILKNKQPALTVVLISSNFINAKLLT
DTIPRYFSDKGIHFYSFYLLRDDIYQIPSLKPDLVITHSRLIPFVKNDLVKGVTVAEFSFDNPDYSIASI
QNLIYQLKDKKYQDFLNEQLQ

SpyoM01000155 is thought to be a collagen binding protein (CPA). An example of SpyoM01000155 is set forth in SEQ ID NO: 244.

SEQ ID NO: 244

MQKRDKTNYGSANNKRRQTTIGLLKVFLTFVALIGIVGFSIRAFGAEEQSVPNRQSSIQDYPWYGYDSYP
KGYPDYSPLKTYHNLKVNLEGSKDYQAYCFNLTKHFPSKSDSVRSQWYKKLEGTNENFIKLADKPRIEDG
QLQQNILRILYNGYPNNRNGIMKGIDPLNAILVTQNAIWYYTDSAQINPDESFKTEARSNGINDQQLGLM
RKALKELIDPNLGSKYSNKTPSGYRLNVFESHDKTFQNLLSAEYVPDTPPKPGEEPPAKTEKTSVIIRKY
AEGDYSKLLEGATLKLSQIEGSGFQEKDFQSNSLGETVELPNGTYTLTETSSPDGYKIAEPIKFRVENKK
VFIVQKDGSQVENPNKEVAEPYSVEAYNDFMDEEVLSGFTPYGKFYYAKNKDKSSQVVYCFNADLHSPPD
SYDSGETINPDTSTMKEVKYTHTAGSDLFKYALRPRDTNPEDFLKHIKKVIEKGYKKKGDSYNGLTETQF
RAATQLAIYYFTDSADLKTLKTYNNGKGYHGFESMDEKTLAVTKELITYAQNGSAPQLTNLDFFVPNNSK
YQSLIGTEYHPDDLVDVIRMEDKKQEVIPVTHSLTVKKTVVGELGDKTKGFQFELELKDKTGQPIVNTLK

TNNQDLVAKDGKYSFNLKHGDTIRIEGLPTGYSYTLKETEAKDYIVTVDNKVSQEAQSVGKDITEDKKVT
FENRKDL*VPPTG*LTTDGAIYLWLLLLVPLGLLVWLFGRKGLKND

SpyoM01000155 contains an amino acid motif indicative of a cell wall anchor: **SEQ ID NO: 184** VPPTG (shown in italics in SEQ ID NO: 244, above). In some recombinant host cell systems, it may be preferable to remove this motif to facilitate secretion of a recombinant SpyoM01000155 protein from the host cell. Alternatively, in other recombinant host cell systems, it may be preferable to use the cell wall anchor motif to anchor the recombinantly expressed protein to the cell wall. The extracellular domain of the expressed protein may be cleaved during purification or the recombinant protein may be left attached to either inactivated host cells or cell membranes in the final composition.

Two pilin motifs, discussed above, containing conserved lysine (K) residues have also been identified in SpyoM01000155. The pilin motif sequence is underlined in SEQ ID NO: 244, below. Conserved lysine (K) residues are also marked in bold, at amino acid residues 71 and 261. The pilin sequences, in particular the conserved lysine residues, are thought to be important for the formation of oligomeric, pilus-like structures. Preferred fragments of SpyoM01000155 include at least one conserved lysine residue. Preferably, fragments include at least one pilin sequence.

SEQ ID NO: 244
MQKRDKTNYGSANNKRRQTTIGLLKVFLTFVALIGIVGFSIRAFGAEEQSVPNRQSSIQDYPWYGYDSYP
KGYPDYSPLKTYHNLKVNLEGSKDYQAYCFNLTKHFPSKSDSVRSQWYKKLEGTNENFIKLADKPRIEDG
QLQQNILRILYNGYPNNRNGIMKGIDPLNAILVTQNAIWYYTDSAQINPDESFKTEARSNGINDQQLGLM
RKALKELIDPNLGSKYSNKTPSGYRLNVFESHDKTFQNLLSAEYVPDTPP**K**PGEEPPAKTEKTSVIIRKY
AEGDYSKLLEGATLKLSQIEGSGFQEKDFQSNSLGETVELPNGTYTLTETSSPDGYKIAEPIKFRVENKK
VFIVQKDGSQVENPNKEVAEPYSVEAYNDFMDEEVLSGFTPYGKFYYAKNKDKSSQVVYCFNADLHSPPD
SYDSGETINPDTSTMKEVKYTHTAGSDLFKYALRPRDTNPEDFLKHIKKVIEKGYKKKGDSYNGLTETQF
RAATQLAIYYFTDSADLKTLKTYNNGKGYHGFESMDEKTLAVTKELITYAQNGSAPQLTNLDFFVPNNSK
YQSLIGTEYHPDDLVDVIRMEDKKQEVIPVTHSLTVKKTVVGELGDKTKGFQFELELKDKTGQPIVNTLK
TNNQDLVAKDGKYSFNLKHGDTIRIEGLPTGYSYTLKETEAKDYIVTVDNKVSQEAQSVGKDITEDKKVT
FENRKDLVPPTGLTTDGAIYLWLLLLVPLGLLVWLFGRKGLKND

Two E boxes containing conserved glutamic residues have been identified in SpyoM01000155. The E-box motifs are underlined in SEQ ID NO: 244, below. The conserved glutamic acid (E) residues, at amino acid residues 329 and 668, are marked in bold. The E box motifs, in particular the conserved glutamic acid residues, are thought to be important for the formation of oligomeric pilus-like structures of SpyoM01000155. Preferred fragments of SpyoM01000155 include at least one conserved glutamic acid residue. Preferably, fragments include at least one E box motif.

SEQ ID NO: 244
MQKRDKTNYGSANNKRRQTTIGLLKVFLTFVALIGIVGFSIRAFGAEEQSVPNRQSSIQDYPWYGYDSYP
KGYPDYSPLKTYHNLKVNLEGSKDYQAYCFNLTKHFPSKSDSVRSQWYKKLEGTNENFIKLADKPRIEDG
QLQQNILRILYNGYPNNRNGIMKGIDPLNAILVTQNAIWYYTDSAQINPDESFKTEARSNGINDQQLGLM
RKALKELIDPNLGSKYSNKTPSGYRLNVFESHDKTFQNLLSAEYVPDTPPKPGEEPPAKTEKTSVIIRKY
AEGDYSKLLEGATLKLSQIEGSGFQEKDFQSNSLGETVELPNGTYTLT**E**TSSPDGYKIAEPIKFRVENKK
VFIVQKDGSQVENPNKEVAEPYSVEAYNDFMDEEVLSGFTPYGKFYYAKNKDKSSQVVYCFNADLHSPPD
SYDSGETINPDTSTMKEVKYTHTAGSDLFKYALRPRDTNPEDFLKHIKKVIEKGYKKKGDSYNGLTETQF

RAATQLAIYYFIDSADLKDLKTYNNGKGYHGFESMDEKTLAVTKELITYAQNGSAPQLTNLDFFVPNNSK
YQSLIGTEYHPDDLVDVIRMEDKKQEVIPVTHSLTVKKTVVGELGDKTKGFQFELELKDKTGQPIVNTLK
TNNQDLVAKDGKYSFNLKHGDTIRIEGLPTGYSYTLKETEAKDYIVTVDNKVSQEAQSVGKDITEDKKVT
FENRKDLVPPTGLTTDGAIYLWLLLLVPLGLLVWLFGRKGLKND

SpyoM01000154 is a LepA protein. An example of SpyoM01000154 is shown in SEQ ID NO: 245.

SEQ ID NO: 245
MTNYLNRLNENSLFKAFIRLVLKISIIGFLGYILFQYVFGVMIINTNDMSPALSAGDGVLYYRLADRSHI
NDVVVYEVDNTLKVGRIAAQAGDEVNFTQEGGLLINGHPPEKEVPYLTYPHSSGPNFPYKVPTGTYFILN
DYREERLDSRYYGALPINQIKGKISTLLRVRGI

SpyoM01000153 is thought to be a fimbrial protein. An example of SpyoM01000153 is shown in SEQ ID NO: 246.

SEQ ID NO: 246
MKKNKLLLATAILATALGMASMSQNIKAETAGVIDGSTLVVKKTFPSYTDDNVLMPKADYSFKVEADDNA
KGKTKDGLDIKPGVIDGLENTKTIRYSNSDKITAKEKSVNFEFANVKFPGVGVYRYTVAEVNGNKAGITY
DSQQWTVDVYVVNKEGGGFEVKYIVSTEVGQSEKKPVLFKNSFDTTSLKIEKQVTGNTGEHQRLFSFTLL
LTPNECFEKGQVVNILQGGETKKVVIGEEYSFTLKDKESVTLSQLPVGIEYKLTEEDVTKDGYKTSATLK
DGEQSSTYELGKDHKTDKSADEIVVTNKRDT*QVPTG*VVGTLAPFAVLSIVAIGGVIYITKRKKA

SpyoM01000153 contains an amino acid motif indicative of a cell wall anchor: **SEQ ID NO: 140** QVPTG (shown in italics in SEQ ID NO: 246, above). In some recombinant host cell systems, it may be preferable to remove this motif to facilitate secretion of a recombinant SpyoM01000153 protein from the host cell. Alternatively, in other recombinant host cell systems, it may be preferable to use the cell wall anchor motif to anchor the recombinantly expressed protein to the cell wall. The extracellular domain of the expressed protein may be cleaved during purification or the recombinant protein may be left attached to either inactivated host cells or cell membranes in the final composition.

A pilin motif, discussed above, containing a conserved lysine (K) residue has also been identified in SpyoM01000153. The pilin motif sequence is underlined in SEQ ID NO: 246, below. A conserved lysine (K) residue is also marked in bold, at amino acid residue 57. The pilin sequence, in particular the conserved lysine residue, is thought to be important for the formation of oligomeric, pilus-like structures. Preferred fragments of SpyoM01000153 include the conserved lysine residue. Preferably, fragments include the pilin sequence.

SEQ ID NO: 246
MKKNKLLLATAILATALGMASMSQNIKAETAGVIDGSTLVVKKTFPSYTDDNVLMPKADYSFKVEADDNA
KGKTKDGLDIKPGVIDGLENTKTIRYSNSDKITAKEKSVNFEFANVKFPGVGVYRYTVAEVNGNKAGITY
DSQQWTVDVYVVNKEGGGFEVKYIVSTEVGQSEKKPVLFKNSFDTTSLKIEKQVTGNTGEHQRLFSFTLL
LTPNECFEKGQVVNILQGGETKKVVIGEEYSFTLKDKESVTLSQLPVGIEYKLTEEDVTKDGYKTSATLK
DGEQSSTYELGKDHKTDKSADEIVVTNKRDTQVPTGVVGTLAPFAVLSIVAIGGVIYITKRKKA

An E box containing a conserved glutamic residue has been identified in SpyoM01000153. The E-box motif is underlined in SEQ ID NO: 246, below. The conserved glutamic acid (E), at amino acid residue 265, is marked in bold. The E box motif, in particular the conserved glutamic acid residue, is thought to be important for the formation of oligomeric pilus-like structures of SpyoM01000153. Preferred fragments of SpyoM01000153 include the conserved glutamic acid residue. Preferably, fragments include the E box motif.

SEQ ID NO: 246
MKKNKLLLATAILATALGMASMSQNIKAETAGVIDGSTLVVKKTFPSYTDDNVLMPKADYSFKVEADDNA
KGKTKDGLDIKPGVIDGLENTKTIRYSNSDKITAKEKSVNFEFANVKFPGVGVYRYTVAEVNGNKAGITY
DSQQWTVDVYVVNKEGGGFEVKYIVSTEVGQSEKKPVLFKNSFDTTSLKIEKQVTGNTGEHQRLFSFTLL
LTPNECFEKGQVVNILQGGETKKVVIGEEYSFTLKDKESVTLSQLPVGIEYKLT**E**EDVTKDGYKTSATLK
DGEQSSTYELGKDHKTDKSADEIVVTNKRDTQVPTGVVGTLAPFAVLSIVAIGGVIYITKRKKA

SpyoM01000152 is a SrtC2 type sortase. An example of SpyoM01000152 is shown in SEQ ID NO: 247

SEQ ID NO: 247
MMMTIVQVINKAIDTLILIFCLVVLFLAGFGLWDSYHLYQQADASNFKKFKTAQQQPKFEDLLALNEDVI
GWLNIPGTHIDYPLVQGKTNLEYINKAVDGSVAMSGSLFLDTRNHNDFTDDYSLIYGHHMAGNAMFGEIP
KFLKKNFFNKHNKAIIETKERKKLTVTIFACLKTDAFDQLVFNPNAITNQDQQRQLVDYISKRSKQFKPV
KLKHHTKFVAFSTCENFSTDNRVIVVGTIQE

SpyoM01000151 is referred to as a hypothetical protein. An example of SpyoM01000151 is shown in SEQ ID NO: 248.

SEQ ID NO: 248
MLFSVVMMLTMLAFNQTVLAKDSTVQTSISVENVLERAGDSTPFSIALESIDAMKTIEEITIAGSGKASF
SPLTFTTVGQYTYRVYQKPSQNKDYQADTTVFDVLVYVTYDEDGTLVAKVISRRAGDEEKSAITFKPKRL
VKPIPPRQPDIPKTP*LPLAG*EVKSLLGILSIVLLGLLVLLYVKKLKSRL

SpyoM01000151 contains an amino acid motif indicative of a cell wall anchor: **SEQ ID NO: 185** LPLAG (shown in italics in SEQ ID NO: 248, above). In some recombinant host cell systems, it may be preferable to remove this motif to facilitate secretion of a recombinant SpyoM01000151 protein from the host cell. Alternatively, in other recombinant host cell systems, it may be preferable to use the cell wall anchor motif to anchor the recombinantly expressed protein to the cell wall. The extracellular domain of the expressed protein may be cleaved during purification or the recombinant protein may be left attached to either inactivated host cells or cell membranes in the final composition.

A pilin motif, discussed above, containing a conserved lysine (K) residue has also been identified in SpyoM01000151. The pilin motif sequence is underlined in SEQ ID NO: 248, below. Conserved lysine (K) residues are also marked in bold, at amino acid residue 138. The pilin sequence, in particular the conserved lysine residue, is thought to be important for the formation of oligomeric, pilus-like structures. Preferred fragments of SpyoM01000151 include the conserved lysine residue. Preferably, fragments include the pilin sequence.

SEQ ID NO: 248
MLFSVVMMLTMLAFNQTVLAKDSTVQTSISVENVLERAGDSTPFSIALESIDAMKTIEEITIAGSGKASF
SPLTFTTVGQYTYRVYQKPSQNKDYQADTTVFDVLVYVTYDEDGTLVAKVISRRAGDEEKSAITFKP**K**RL
VKPIPPRQPDIPKTPLPLAGEVKSLLGILSIVLLGLLVLLYVKKLKSRL

Two E boxes containing conserved glutamic residues have been identified in SpyoM01000151. The E-box motifs are underlined in SEQ ID NO: 248, below. The conserved glutamic acid (E) residues, at amino acid residues 58 and 128, are marked in bold. The E box motifs, in particular the conserved glutamic acid residues, are thought to be important for the formation of oligomeric pilus-like structures of SpyoM01000151. Preferred fragments of SpyoM01000151 include at least one conserved glutamic acid residue. Preferably, fragments include at least one E box motif.

SEQ ID NO: 248
MLFSVVMMLTMLAFNQTVLAKDSTVQTSISVENVLERAGDSTPFSIALESIDAMKTIEEITIAGSGKASF
SPLTFTTVGQYTYRVYQKPSQNKDYQADTTVFDVLVYVTYDEDGTLVAKVISRRAGDEEKSAITFKPKRL
VKPIPPRQPDIPKTPLPLAGEVKSLLGILSIVLLGLLVLLYVKKLKSRL

SpyoM01000150 is referred to as a putative MsmRL. An example of SpyoM01000150 is set forth in SEQ ID NO: 249.

SEQ ID NO: 249
MVIFDLKHVQTLHSLSQLPISVMSQDKALIQVYGNDDYLLCYYQFLKHLAIPQAAQDVIFYEGLFEESFM
IFPLCHYIIAIGPFYPYSLNKDYQEQLANNFLKHSSHRSKEELLSYMALVPHFPINNVRNLLIAIDAFFD
TQFETTCQQTIHQLLQHSKQMTADPDIIHRLKHISKASSQLPPVLEHLNHIMDLVKLGNPQLLKQEINRI
PLSSITSSSISALRAEKNLTVIYLTRLLEFSFVENTDVAKHYSLVKYYMALNEEASDLLKVLRIRCAAII
HFSESLTNKSISDKRQMYNSVLHYVDSHLYSKLKVSDIAKRLYVSESHLRSVFKKYSNVSLQHYILSTKI
KEAQLLLKRGIPVGEVAKSLYFYDTTHFHKIFKKYTGISSKDYLAKYRDNI

SpyoM01000149 is a F2 like fibronectin-binding protein. An example of SpyoM01000149 is set forth in SEQ ID NO: 250.

SEQ ID NO: 250
MTQKNSYKLSFLLSLTGFILGLLLVFIGLSGVSVGHAETRNGANKQGYFEIKKVDQNNKPLSGATFSLTP
KDGKGKPVQTFTSSEEGIIDAQNLQPGTYTLKEETAPDGYDKTSRTWTVTVYENGYTKLVENPYNGEIIS
KAGSKDVSSSLQLENPKMSVVSKYGEQEKTSNSADFYRNHAAYFKMSFELKQKDKSETINPGDTFVLQLD
RRLNPKGISQDIPKIIYDSENSPLAIGKYDAKTHQLTYTFTNYIAGLDKVQLSAELSLFLENKEVLENTN
ISDFKSTIGGQEITYKGTVNVLYGNESTKESNYITNGLSNVGGSIESYNTETGEFVWYVYVNPNRTNIPY
AVLNLWGFAKRTAQGENDNSSVSSAQLTGYDIYEVPHNYRLPTSYGVDISRLNLRKDLEAKLPQGSTQGA
NKRLRIDFGENLQGKAFVVKVTGKADQSGKELIVQSHLSSFNNWGSYKTLRPNSHVSFTNEIALSPSKGS
GSGTSEFTKPAITVANLKRVAQLRFKKVSTDNVPLPEAAFELRSSNGNSQKLEASSNTQGEIHFKDLTSG
TYDLYETKAPKGYQQVTEKLATVTVDTTKPAEQMVKWEKPHSFVKVEANKEVTIVNHKETLTFSGKKIWE
NDRPDQRPAKIQVQLLQNGQKMPNQIQEVTKDNDWSYHFKDLPKYDAKNQEYKYSVEEVKVPDGYKVSYL
GNDIFNTRETEFVFEQNNFNLEFGNAEIKGQSGSKIIDEDTLTSFKGKKIWKNDTAENRPQAIQVQLYAD
GVAVEGQTKFISGSGNEWSFEFKNLKKYNGTGNDIIYSVKEVTVPTGYDVTYSANDIINTKREVITQQGP
NLEIEETLPLESGASGGTTTVEDSRSVDTLSGLSSEQGQSGDMTIEEDSATHIKFSKRDIDGKELAGATM
ELRDSSGKTISTWISDGQVKDFYLMPGKYTFVETAAPDGYEIATAITFTVNEQGQVTVNGKATKGDAHIV
MVDAYKPTKGSGQVIDIEEKLPDEQGHSGSTTEIEDSKPSDVIIGGQGEVVDTTEDTQSGMTGHSGSTTE
IEDSKSSDVIIGGQGQVVETTEDTQTGMHGDSGCKTEVEDTKLVQFFHFDNKEPESNSEIPKKDKPKSNT
S*LPATG*EKQHNKFFWMVTSCSLISSVFVISLKSKKRLLSC

SpyoM01000149 contains an amino acid motif indicative of a cell wall anchor: **SEQ ID NO: 180** LPATG (shown in italics in SEQ ID NO: 250, above). In some recombinant host cell systems, it may be preferable to remove this motif to facilitate secretion of a recombinant SpyoM01000149 protein from the host cell. Alternatively, in other recombinant host cell systems, it may be preferable to use the cell wall anchor motif to anchor the recombinantly expressed protein to the cell wall. The extracellular domain of the expressed protein may be cleaved during purification or the recombinant protein may be left attached to either inactivated host cells or cell membranes in the final composition.

Two pilin motifs, discussed above, containing conserved lysine (K) residues have also been identified in SpyoM01000149. The pilin motif sequences are underlined in SEQ ID NO: 250, below. Conserved lysine (K) residues are also marked in bold, at amino acid residues 157 and 163, and 216 and 224. The pilin sequences, in particular the conserved lysine residues, are thought to be important for the formation of oligomeric, pilus-like structures. Preferred fragments of SpyoM01000149 include at least one conserved lysine residue. Preferably, fragments include at least one pilin sequence.

SEQ ID NO: 250

MTQKNSYKLSFLLSLTGFILGLLLVFIGLSGVSVGHAETRNGANKQGYFEIKKVDQNNKPLSGATFSLTP
KDGKGKPVQTFTSSEEGIIDAQNLQPGTYTLKEETAPDGYDKTSRTWTVTVYENGYTKLVENPYNGEIIS
KAGSKDVSSSLQLENP**K**MSVVS**K**YGEQEKTSNSADFYRNHAAYFKMSFELKQKDKSETINPGDTFVLQLD
RRLNP**K**GISQDIP**K**IIYDSENSPLAIGKYDAKTHQLTYTFTNYIAGLDKVQLSAELSLFLENKEVLENTN
ISDFKSTIGGQEITYKGTVNVLYGNESTKESNYITNGLSNVGGSIESYNTETGEFVWYVYVNPNRTNIPY
AVLNLWGFAKRTAQGENDNSSVSSAQLTGYDIYEVPHNYRLPTSYGVDISRLNLRKDLEAKLPQGSTQGA
NKRLRIDFGENLQGKAFVVKVTGKADQSGKELIVQSHLSSFNNWGSYKTLRPNSHVSFTNEIALSPSKGS
GSGTSEFTKPAITVANLKRVAQLRFKKVSTDNVPLPEAAFELRSSNGNSQKLEASSNTQGEIHFKDLTSG
TYDLYETKAPKGYQQVTEKLATVTVDTTKPAEQMVKWEKPHSFVKVEANKEVTIVNHKETLTFSGKKIWE
NDRPDQRPAKIQVQLLQNGQKMPNQIQEVTKDNDWSYHFKDLPKYDAKNQEYKYSVEEVKVPDGYKVSYL
GNDIFNTRETEFVFEQNNFNLEFGNAEIKGQSGSKIIDEDTLTSFKGKKIWKNDTAENRPQAIQVQLYAD
GVAVEGQTKFISGSGNEWSFEFKNLKKYNGTGNDIIYSVKEVTVPTGYDVTYSANDIINTKREVITQQGP
NLEIEETLPLESGASGGTTTVEDSRSVDTLSGLSSEQGQSGDMTIEEDSATHIKFSKRDIDGKELAGATM
ELRDSSGKTISTWISDGQVKDFYLMPGKYTFVETAAPDGYEIATAITFTVNEQGQVTVNGKATKGDAHIV
MVDAYKPTKGSGQVIDIEEKLPDEQGHSGSTTEIEDSKPSDVIIGGQGEVVDTTEDTQSGMTGHSGSTTE
IEDSKSSDVIIGGQGQVVETTEDTQTGMHGDSGCKTEVEDTKLVQFFHFDNKEPESNSEIPKKDKPKSNT
SLPATGEKQHNKFFWMVTSCSLISSVFVISLKSKKRLLSC

Two E boxes containing conserved glutamic residues have been identified in SpyoM01000149. The E-box motifs are underlined in SEQ ID NO: 250, below. The conserved glutamic acid (E) residues, at amino acid residues 329 and 668, are marked in bold. The E box motifs, in particular the conserved glutamic acid residues, are thought to be important for the formation of oligomeric pilus-like structures of SpyoM01000149. Preferred fragments of SpyoM01000149 include at least one conserved glutamic acid residue. Preferably, fragments include at least one E box motif.

SEQ ID NO: 250

MTQKNSYKLSFLLSLTGFILGLLLVFIGLSGVSVGHAETRNGANKQGYFEIKKVDQNNKPLSGATFSLTP
KDGKGKPVQTFTSSEEGIIDAQNLQPGTYTLKEETAPDGYDKTSRTWTVTVYENGYTKLVENPYNGEIIS
KAGSKDVSSSLQLENPKMSVVSKYGEQEKTSNSADFYRNHAAYFKMSFELKQKDKSETINPGDTFVLQLD
RRLNPKGISQDIPKIIYDSENSPLAIGKYDAKTHQLTYTFTNYIAGLDKVQLSAELSLFLENKEVLENTN
ISDFKSTIGGQEITYKGTVNVLYGNESTKESNYITNGLSNVGGSIESYNTETGEFVWYVYVNPNRTNIPY
AVLNLWGFAKRTAQGENDNSSVSSAQLTGYDIYEVPHNYRLPTSYGVDISRLNLRKDLEAKLPQGSTQGA
NKRLRIDFGENLQGKAFVVKVTGKADQSGKELIVQSHLSSFNNWGSYKTLRPNSHVSFTNEIALSPSKGS
GSGTSEFTKPAITVANLKRVAQLRFKKVSTDNVPLPEAAFELRSSNGNSQKLEASSNTQGEIHFKDLTSG
TYDLY**E**TKAPKGYQQVTEKLATVTVDTTKPAEQMVKWEKPHSFVKVEANKEVTIVNHKETLTFSGKKIWE
NDRPDQRPAKIQVQLLQNGQKMPNQIQEVTKDNDWSYHFKDLPKYDAKNQEYKYSVEEVKVPDGYKVSYL
GNDIFNTRETEFVFEQNNFNLEFGNAEIKGQSGSKIIDEDTLTSFKGKKIWKNDTAENRPQAIQVQLYAD
GVAVEGQTKFISGSGNEWSFEFKNLKKYNGTGNDIIYSVKEVTVPTGYDVTYSANDIINTKREVITQQGP
NLEIEETLPLESGASGGTTTVEDSRSVDTLSGLSSEQGQSGDMTIEEDSATHIKFSKRDIDGKELAGATM
ELRDSSGKTISTWISDGQVKDFYLMPGKYTFV**E**TAAPDGYEIATAITFTVNEQGQVTVNGKATKGDAHIV
MVDAYKPTKGSGQVIDIEEKLPDEQGHSGSTTEIEDSKPSDVIIGGQGEVVDTTEDTQSGMTGHSGSTTE
IEDSKSSDVIIGGQGQVVETTEDTQTGMHGDSGCKTEVEDTKLVQFFHFDNKEPESNSEIPKKDKPKSNT
SLPATGEKQHNKFFWMVTSCSLISSVFVISLKSKKRLLSC

As discussed above, applicants have also determined the nucleotide and encoded amino acid sequence of fimbrial structural subunits in several other GAS AI-3 strains of bacteria. Examples of sequences of these fimbrial structural subunits are set forth below.

M3 strain isolate ISS 3040 is a GAS AI-3 strain of bacteria. ISS3040_fimbrial is thought to be a fimbrial structural subunit of M3 strain isolate ISS 3040. An example of a nucleotide sequence encoding the ISS3040_fimbrial protein (SEQ ID NO: 263) and an ISS3040_fimbrial protein amino acid sequence (SEQ ID NO: 264) are set forth below.

SEQ ID NO: 263

gagacggcaggagtgtccgaaaatgcaaaattaatagtaaaaaagacatttgactcttat
acagacaatgaagttttaatgccaaaagctgattatacttttaaagtagaggcagatagt
acagctagtggcaaaacgaaagacggtttagagattaagccaggtattgttaatggttta
acagaacagattatcagctatactaatactgataaaccagatagtaaagttaaaagtaca
gagtttgattttcaaaagtagtattccctggtattggtgtttaccgctatactgtttca
gaaaaacaaggtgatgttgaaggaattacctacgatactaagaagtggacagtagatgtt
tatgttggaaacaaagaaggtggtggttttgaacctaagtttattgtatctaaggaacaa
ggaacagacgtcaaaaaaccagttaattttaacaactcgtttgcaactacttcgttaaaa
gttaagaagaatgtatcggggaatactggagaattgcaaaaagaatttgactttacattg
acgcttaatgaaagcacgaattttaaaaaagatcaaattgtttctttacaaaaaggaaac
gagaaatttgaagttaagattggtactccctacaagtttaaactcaaaaatggggaatct
attcaactagacaagttaccagttggtattacttataaagtcaatgaaatggaagctaat
aaagatgggtataaaacaacagcatccttgaaagagggagatggtcaatctaaaatgtat
caattggatatggaacaaaaaacagacgaatctgctgacgaaatcgttgtcacaaataag
cgtgacactcaagttccaactggtgttgtaggcacccttgctccatttgcagttcttagc

SEQ ID NO: 264

ETAGVSENAKLIVKKTFDSYTDNEVLMPKADYTFKVEADSTASG
KTKDGLEIKPGIVNGLTEQIISYTNTDKPDSKVKSTEFDFSKVVFPGIGVYRYTVSEK
QGDVEGITYDTKKWTVDVYVGNKEGGGFEPKFIVSKEQGTDVKKPVNFNNSFATTSLK
VKKNVSGNTGELQKEFDFTLTLNESTNFKKDQIVSLQKGNEKFEVKIGTPYKFKLKNG
ESIQLDKLPVGITYKVNEMEANKDGYKTTASLKEGDGQSKMYQLDMEQKTDESADEIV
VTNKRDTQVPTGVVGTLAPFAVLS

M44 strain isolate ISS 3776 is a GAS AI-3 strain of bacteria. ISS3776_fimbrial is thought to be a fimbrial structural subunit of M44 isolate ISS 3776. An example of a nucleotide sequence encoding the ISS3776_fimbrial protein (SEQ ID NO: 253) and an ISS3776_fimbrial protein amino acid sequence (SEQ ID NO: 254) are set forth below.

SEQ ID NO: 253

ttggagagagaaaaaatgaaaaaaaacaaattattacttgctactgcaatcttagcaact
gctttaggaacagcttctttaaatcaaaacgtaaaagctgagacggcaggggttgtaaca
ggaaaatcactacaagttacaaagacaatgacttatgatgatgaagaggtgttaatgccc
gaaaccgcctttacttttactatagagcctgatatgactgcaagtggaaaagaaggcagc
ctagatattaaaaatggaattgtagaaggcttagacaaacaagtaacagtaaaatataag
aatacagataaaacatctcaaaaaactaaaatagcacaatttgatttttctaaggttaaa
tttccagctataggtgtttaccgctatatggtttcagagaaaaacgataaaaaagacgga
attacgtacgatgataaaaagtggactgtagatgtttatgttgggaataaggccaataac
gaagaaggtttcgaagttctatatattgtatcaaaagaaggtacttctagtactaaaaaa
ccaattgaatttacaaactctattaaaactacttccttaaaaattgaaaaacaaataact
ggcaatgcaggagatcgtaaaaaatcattcaacttcacattaacattacaaccaagtgaa
tattataaaactggatcagttgtgaaaatcgaacaggatggaagtaaaaaagatgtgacg
ataggaacgccttacaaatttactttgggacacggtaagagtgtcatgttatcgaaatta
ccaattggtatcaattactatcttagtgaagacgaagcgaataaagacggctacactaca
acggcaacattaaaagaacaaggcaaagaaaagagttccgatttcactttgagtactcaa
aaccagaaaacagacgaatctgctgacgaaatcgttgtcacaaataagcgtgacactcaa
gttccaactggtgttgtagggacccttgctccatttgcagttcttagcattgtggctatt
ggtggagttatctatattacaaaacgtaaaaaagcttaa

SEQ ID NO: 254

MEREKMKKNKLLLATAILATALGTASLNQNVKAETAGVVTGKSL
QVTKTMTYDDEEVLMPETAFTFTIEPDMTASGKEGSLDIKNGIVEGLDKQVTVKYKNT
DKTSQKTKIAQFDFSKVKFPAIGVYRYMVSEKNDKKDGITYDDKKWTVDVYVGNKANN
EEGFEVLYIVSKEGTSSTKKPIEFTNSIKTTSLKIEKQITGNAGDRKKSFNFTLTLQP
SEYYKTGSVVKIEQDGSKKDVTIGTPYKFTLGHGKSVMLSKLPIGINYYLSEDEANKD

GYTTTATLKEQGKEKSSDFTLSTQNQKTDESADEIVVTNKRDTQVPTGVVGTLAPFAV
LSIVAIGGVIYITKRKKA

M77 strain isolate ISS4959 is a GAS AI-3 strain of bacteria. ISS4959_fimbrial is thought to be a fimbrial structural subunit of M77 strain ISS 4959. An example of a nucleotide sequence encoding the ISS4959_fimbrial protein (SEQ ID NO: 271) and an ISS4959_fimbrial protein amino acid sequence (SEQ ID NO: 272) are set forth below.

SEQ ID NO: 271

gtaacagtaaaatataagaatacagataaaacatctcaaaaaactaaaatagcacaattt
gatttttctaaggttaaatttccagctataggtgtttaccgctatatggtttcagagaaa
aacgataaaaaagacggaattacgtacgatgataaaaagtggacngtagatgtttatgtt
gggaataaggccaataacgaagaaggtttcgaagttctatatattgtatcaaaagaaggt
acttctagtnctaaaaaaccaattgaatttacaaactctattaaaactacttccttaaaa
attgaaaaacaaataactggcaatgcaggagatcgtaaaaaatcattcaacttcacattn
acattacanccaagtgaatattataaaactggatcagttgtgaaaatcgaacaggatgga
agtaaaaaagatgtgacgataggaacgccttacaaatttactttgggacacggtaagagt
gtcatgttatcgaaattnccaattggtatcaattactatcttagtgaagacgaagcgaat
aaagacggntacactacancggcaacattaaaagaacaaggcaaagaaaagagttccgat
ttcactttgagtactcaaaaccagaaaacagacgaatctgctg

SEQ ID NO: 272

VTVKYKNTDKTSQKTKIAQFDFSKVKFPAIGVYRYMVSEKNDKK
DGITYDDKKWTVDVYVGNKANNEEGFEVLYIVSKEGTSSXKKPIEFTNSIKTTSLKIE
KQITGNAGDRKKSFNFTXTLXPSEYYKTGSVVKIEQDGSKKDVTIGTPYKFTLGHGKS
VMLSKXPIGINYYLSEDEANKDGYTTXATLKEQGKEKSSDFTLSTQNQKTDESA

Examples of GAS AI-4 sequences from M12 strain isolate A735 are set forth below.

19224133 is thought to be a RofA regulatory protein. An example of a nucleotide sequence encoding the RofA regulatory protein (SEQ ID NO: 104) and a RofA regulatory protein amino acid sequence (SEQ ID NO: 105) are set forth below.

SEQ ID NO: 104

ATGACCATCCAAAAAAGGATGATATCTTGCCAATTTACACATCCTTCTAAAGAAACTTATCTTTACCAACTCTAT
GCATCATCTAATGTCTTACAATTACTAGCGTTTTTAATAAAAAATGGTTCCCACTCTCGTCCCCTTACGGATTTT
GCAAGAAGTCATTTTTTATCAAACTCCTCAGCTTATCGGATGCGCGAAGCATTGATTCCTTTATTAAGAAACTTT
GAATTAAAACTCTCTAAGAACAAGATTGTCGGTGAGGAATATCGTATCCGTTACCTCATCGCTCTGCTATATAGT
AAGTTTGGCATTAAAGTTTATGACTTGACGCAGCAAGACAAAAACATTATTCATAGCTTTTTATCCCATAGTTCC
ACCCACCTTAAAACTTCTCCTTGGTTATCGGAATCGTTTTCTTTCTATGACATTTTATTAGCTTTATCGTGGAAG
CGGCATCAATTTTCGGTAACTATTCCCCAAACCAGAATTTTTCAACAATTAAAAAAACTTTTTGTCTACGATTCT
TTGAAAAAAAGTAGCCGTGATATTATCGAAACTTACTGCCAACTAAACTTTTCAGCAGGAGATTTGGACTACCTC
TATTTAATTTATATCACCGCTAATAATTCTTTTGCGAGCTTACAATGGACACCTGAGCATATCAGACAATGTTGT
CAACTTTTTGAAGAAAATGATACTTTTCGCCTGCTTTTAAATCCTATCATCACTCTTTTACCTAACCTAAAAGAG
CAAAAGGCTAGTTTAGTAAAAGCTCTTATGTTTTTTTCAAAATCATTCTTGTTTAATCTGCAACATTTTATTCCT
GAGACCAACTTATTCGTTTCTCCGTACTATAAAGGAAACCAAAAACTCTATACGTCCTTAAAGTTAATTGTCGAA
GAGTGGATGGCCAAACTTCCTGGTAAGCGTTACTTGAACCATAAGCATTTTCATCTTTTTTGCCACTATGTCGAG
CAAATTCTAAGAAATATCCAACCTCCTTTAGTTGTTGTTTTCGTAGCCAGTAATTTTATCAATGCTCATCTCCTA
ACAGATTCTTTCCCAAGGTATTTCTCGGATAAAAGCATTGATTTTCATTCCTATTATCTATTGCAAGATAATGTT
TATCAAATTCCTGATTTAAAGCCAGATTTGGTCATCACTCACAGTCAACTGATTCCTTTTGTTCACCATGAACTT
ACAAAAGGAATTGCTGTTGCTGAAATATCTTTTGATGAATCGATTCTGTCTATCCAAGAATTGATGTATCAAGTT
AAAGAGGAAAAATTCCAAGCTGATTTAACCAAACAATTAACATAA

SEQ ID NO: 105

MTIQKRMISCQFTHPSKETYLYQLYASSNVLQLLAFLIKNGSHSRPLTDFARSHFLSNSSAYRMREALIPLLRNF
ELKLSKNKIVGEEYRIRYLIALLYSKFGIKVYDLTQQDKNIIHSFLSHSSTHLKTSPWLSESFSFYDILLALSWK
RHQFSVTIPQTRIFQQLKKLFVYDSLKKSSRDIIETYCQLNFSAGDLDYLYLIYITANNSFASLQWTPEHIRQCC
QLFEENDTFRLLLNPIITLLPNLKEQKASLVKALMFFSKSFLFNLQHFIPETNLFVSPYYKGNQKLYTSLKLIVE

EWMAKLPGKRYINNHKHFHLFCHYVEQILRNIQPPLVVVFVASNFINAHLLTDSFPRYFSDKSIDFHSYYLLQDNV
YQIPDLKPDLVITHSQLIPFVHHELTKGIAVAEISFDESILSIQELMYQVKEEKFQADLTKQLT

19224134 is thought to be a protein F fibronectin binding protein. An example of a nucleotide sequence encoding the protein F fibronectin binding protein (SEQ ID NO: 106) and a protein F fibronectin binding protein amino acid sequence (SEQ ID NO: 107) are set forth below.

SEQ ID NO: 106

ATGGTAAGCTCATATATGTTTGCGAGAGGAGAGAAAATGAATAACAAAATGTTTTTGAACAAAGAAGCCGGTTTT
TTGGTACACACAAAAAGAAAAAGGCGATTTGCTGTCACTTTAGTGGGAGTCTTTTTTCTGCTTTTGGCATGTGCG
GGTGCTATCGGTTTTGGTCAAGTAGCCTATGCTGCGGATGAGAAGACTGTGCCGAATTTTAAAAGCCCAGATCCA
GATTATCCCTGGTATGGTTATGATTCGTATAGAGGAATATTTGCAAGATATCACAATTTAAAAGTAAATCTAAAA
GGAAGTAAGGAGTATCAAGCGTATTGTTTTAACCTAACAAAATACTTTCCTCGCCCCACTTATAGTACTACAAAT
AATTTTTACAAGAAAATTGATGGGAGTGGATCAGCGTTCAAATCTTATGCAGCGAATCCTAGGGTTTTAGATGAG
AATTTAGATAAATTAGAAAAAAATATACTGAATGTAATTTATAATGGATATAAAAGTAATGCAAATGGTTTTATG
AATGGTATAGAAGATCTTAATGCTATACTAGTAACTCAAAACGCTATTTGGTACTATTCAGATAGTGCTCCATTA
AATGATGTTAATAAAATGTGGGAAAGAGAGGTTCGGAATGGGGAGATTAGTGAGTCACAAGTTACTTTAATGCGT
GAGGCATTGAAAAAACTAATTGATCCCAATTTAGAAGCTACTGCAGCTAATAAAATCCCATCAGGATATCGTTTA
AATATCTTTAAGTCTGAAAATGAAGATTACCAAAATCTTTTAAGTGCTGAATATGTACCTGATGATCCCCCTAAA
CCTGGTGATACGTCAGAACATAATCCTAAAACTCCCGAGTTGGATGGCACTCCAATTCCCGAGGACCCAAAACGT
CCAGATGAGAGTTCAGAACCTGCGCTTCCCCCATTAATGCCAGAGCTAGATGGTGAAGAAGTCCCAGAAGTTCCA
AGCGAGAGCTTAGAACCTGCGCTTCCCCCATTGATGCCAGAGCTAGATGGTGAAGAAGTCCCAGAAGTTCCAAGC
GAGAGCTTAGAACCTGCGCTTCCCCCATTGATGCCAGAGCTAGATGGTGAAGAAGTCCCAGAAGTTCCAAGCGAG
AGCTTAGAACCTGCGCTTCCCCCATTAATGCCAGAGCTAGATGGTGAAGAAGTCCCAGAAGTTCCAAGCGAGAGC
TTAGAACCTGCGCTTCCCCCATTGATGCCAGAGTTAGATGGTGAAGAAGTCCCTGAAAAACCTAGTGTTGACTTA
CCTATTGAAGTTCCTCGTTATGAGTTTAACAATAAAGACCAGTCACCTCTAGCGGGTGAGTCTGGTGAGACGGAG
TATATTACCGAAGTCTATGGAAATCAACAGAACCCTGTTGATATTGATAAAAAACTTCCGAATGAAACAGGTTTT
TCAGGAAATATGGTTGAGACAGAAGATACGAAAGAGCCAGAAGTGTTGATGGGAGGTCAAAGTGAGTCTGTTGAA
TTTACTAAAGACACTCAAACAGGCATGAGTGGTCAAACAACTCCTCAGGTTGAGACAGAAGATACGAAAGAGCCA
GAAGTGTTGATGGGAGGTCAAAGTGAGTCTGTTGAATTTACTAAAGACACTCAAACAGGCATGAGTGGTCAAACA
ACTCCTCAGGTTGAGACAGAAGATACGAAAGAGCCAGGAGTGTTGATGGGAGGCCAAAGTGAGTCTGTTGAATTT
ACTAAAGACACTCAAACAGGCATGAGTGGTCAAACAACTCCTCAGGTTGAGACAGAAGACACGAAAGAGCCAGGA
GTGTTGATGGGAGGTCAAAGTGAGTCTGTTGAATTTACTAAAGACACTCAAACAGGCATGAGCGGTTTCAGTGAA
ACAGTGACCATTGTTGAAGATACGCGTCCGAAGTTAGTGTTCCATTTTGACAATAATGAGCCCAAAGTGGAAGAG
AATCGGGAAAAGCCTACAAAAAATATAACACCTATCCTTCCTGCAACAGGAGATATTGAGAATGTTTTGGCCTTT
CTTGGAATCCTTATTTTGTCAGTACTTTCTATTTTTAGCCTTTTAAAAAACAAACAAAACAATAAAGTCTGA

SEQ ID NO: 107

MVSSYMFARGEKMNNKMFLNKEAGFLVHTKRKRRFAVTLVGVFFLLLACAGAIGFGQVAYAADEKTVPNFKSPDP
DYPWYGYDSYRGIFARYHNLKVNLKGSKEYQAYCFNLTKYFPRPTYSTTNNFYKKIDGSGSAFKSYAANPRVLDE
NLDKLEKNILNVIYNGYKSNANGFMNGIEDLNAILVTQNAIWYYSDSAPLNDVNKMWEREVRNGEISESQVTLMR
EALKKLIDPNLEATAANKIPSGYRLNIFKSENEDYQNLLSAEYVPDDPPKPGDTSEHNPKTPELDGTPIPEDPKR
PDESSEPALPPLMPELDGEEVPEVPSESLEPALPPLMPELDGEEVPEVPSESLEPALPPLMPELDGEEVPEVPSE
SLEPALPPLMPELDGEEVPEVPSESLEPALPPLMPELDGEEVPEKPSVDLPIEVPRYEFNNKDQSPLAGESGETE
YITEVYGNQQNPVDIDKKLPNETGFSGNMVETEDTKEPEVLMGGQSESVEFTKDTQTGMSGQTTPQVETEDTKEP
EVLMGGQSESVEFTKDTQTGMSGQTTPQVETEDTKEPGVLMGGQSESVEFTKDTQTGMSGQTTPQVETEDTKEPG
VLMGGQSESVEFTKDTQTGMSGFSETVTIVEDTRPKLVFHFDNNEPKVEENREKPTKNITPI*LPATG*DIENVLAF
LGILILSVLSIFSLLKNKQNNKV

19224134 contains an amino acid motif indicative of a cell wall anchor: **SEQ ID NO: 181** LPATG (shown in italics in SEQ ID NO: 107, above). In some recombinant host cell systems, it may be preferable to remove this motif to facilitate secretion of a recombinant 19224134 protein from the host cell. Alternatively, in other recombinant host cell systems, it may be preferable to use the cell wall anchor motif to anchor the recombinantly expressed protein to the cell wall. The extracellular domain of the expressed protein may be cleaved during purification or the recombinant protein may be left attached to either inactivated host cells or cell membranes in the final composition.

A pilin motif, discussed above, containing a conserved lysine (K) residue has also been identified in 19224134. The pilin motif sequence is underlined in SEQ ID NO: 107, below. Conserved lysine (K) residues are also marked in bold, at amino acid residues 275, 285, and 299. The pilin sequence, in particular the conserved lysine residues, are thought to be important for the formation of oligomeric, pilus-like structures. Preferred fragments of 19224134 include at least one conserved lysine residue. Preferably, fragments include the pilin sequence.

SEQ ID NO: 107

MVSSYMFARGEKMNNKMFLNKEAGFLVHTKRKRRFAVTLVGVFFLLLACAGAIGFGQVAYAADEKTVPNFKSPDP
DYPWYGYDSYRGIFARYHNLKVNLKGSKEYQAYCFNLTKYFPRPTYSTTNNFYKKIDGSGSAFKSYAANPRVLDE
NLDKLEKNILNVIYNGYKSNANGFMNGIEDLNAILVTQNAIWYYSDSAPLNDVNKMWEREVRNGEISESQVTLMR
EALKKLIDPNLEATAANKIPSGYRLNIFKSENEDYQNLLSAEYVPDDPP**K**PGDTSEHNP**K**TPELDGTPIPEDP**K**R
PDESSEPALPPLMPELDGEEVPEVPSESLEPALPPLMPELDGEEVPEVPSESLEPALPPLMPELDGEEVPEVPSE
SLEPALPPLMPELDGEEVPEVPSESLEPALPPLMPELDGEEVPEKPSVDLPIEVPRYEFNNKDQSPLAGESGETE
YITEVYGNQQNPVDIDKKLPNETGFSGNMVETEDTKEPEVLMGGQSESVEFTKDTQTGMSGQTTPQVETEDTKEP
EVLMGGQSESVEFTKDTQTGMSGQTTPQVETEDTKEPGVLMGGQSESVEFTKDTQTGMSGQTTPQVETEDTKEPG
VLMGGQSESVEFTKDTQTGMSGFSETVTIVEDTRPKLVFHFDNNEPKVEENREKPTKNITPILPATGDIENVLAF
LGILILSVLSIFSLLKNKQNNKV

Two E boxes containing conserved glutamic residues have been identified in 19224134. The E-box motifs are underlined in SEQ ID NO: 107, below. The conserved glutamic acid (E) residues, at amino acid residues 487 and 524, are marked in bold. The E box motifs, in particular the conserved glutamic acid residues, are thought to be important for the formation of oligomeric pilus-like structures of 19224134. Preferred fragments of 19224134 include at least one conserved glutamic acid residue. Preferably, fragments include at least one E box motif.

SEQ ID NO: 107

MVSSYMFARGEKMNNKMFLNKEAGFLVHTKRKRRFAVTLVGVFFLLLACAGAIGFGQVAYAADEKTVPNFKSPDP
DYPWYGYDSYRGIFARYHNLKVNLKGSKEYQAYCFNLTKYFPRPTYSTTNNFYKKIDGSGSAFKSYAANPRVLDE
NLDKLEKNILNVIYNGYKSNANGFMNGIEDLNAILVTQNAIWYYSDSAPLNDVNKMWEREVRNGEISESQVTLMR
EALKKLIDPNLEATAANKIPSGYRLNIFKSENEDYQNLLSAEYVPDDPPKPGDTSEHNPKTPELDGTPIPEDPKR
PDESSEPALPPLMPELDGEEVPEVPSESLEPALPPLMPELDGEEVPEVPSESLEPALPPLMPELDGEEVPEVPSE
SLEPALPPLMPELDGEEVPEVPSESLEPALPPLMPELDGEEVPEKPSVDLPIEVPRYEFNNKDQSPLAGESGETE
YITEVYGNQQNPVDIDKKLPNETGFSGNMVETEDTK**E**PEVLMGGQSESVEFTKDTQTGMSGQTTPQVETEDTK**E**P
EVLMGGQSESVEFTKDTQTGMSGQTTPQVETEDTKEPGVLMGGQSESVEFTKDTQTGMSGQTTPQVETEDTKEPG
VLMGGQSESVEFTKDTQTGMSGFSETVTIVEDTRPKLVFHFDNNEPKVEENREKPTKNITPILPATGDIENVLAF
LGILILSVLSIFSLLKNKQNNKV

19224135 is thought to be a capsular polysaccharide adhesin (Cpa) protein. An example of a nucleotide sequence encoding the Cpa protein (SEQ ID NO: 108) and a Cpa protein amino acid sequence (SEQ ID NO: 109) are set forth below.

SEQ ID NO: 108

ATGAATAACAAAAAATTGCAAAAGAAGCAAGATGCTCCTCGGGTATCAAACAGAAAGCCAAAACAATTAACTGTC
ACTTTAGTGGGAGTATTTTTAATGTTTTTGACCTTGGTAAGTTCCATGAGAGGTGCTCAAAGCATATTTGGAGAG
GAAAAGAGAATTGAAGAAGTCAGTGTTCCTAAAATAAAAAGTCCAGATGATGCCTACCCTTGGTATGGCTATGAT
TCATATGACTCTAGTCATCCTTACTATGAACGTTTTAAAGTAGCACATGATTTAAGGGTTAATTTAAATGGAAGT
AAGAGCTACCAAGTATATTGCTTTAATATCAATTCTCATTATCCGAATAGAAAAAATGCTTTTTCTAAACAATGG
TTTAAGAGAGTTGATGGGACAGGTGATGTGTTCACAAATTATGCTCAGACACCTAAGATTCGTGGAGAATCATTG
AATAATAAACTTTTAAGTATTATGTACAACGCTTATCCTAAAAATGCTAATGGCTATATGGATAAGATAGAACCA

TTAAATGCTATTTTAGTAACTCAACAAGCTGTTTGGTACTATTCTGACAGTTCTTATGGTAATATAAAAACGTTA
TGGGCATCTGAGCTTAAAGACGGAAAAATAGATTTTGAACAAGTAAAATTAATGCGTGAAGCTTACTCAAAACTA
ATTAGTGATGATTTAGAAGAAACATCTAAAAATAAGCTACCTCAAGGATCTAAACTGAATATTTTTGTTCCGCAA
GATAAATCTGTTCAAAATTTATTAAGTGCAGAGTACGTGCCTGAATCCCCTCCGGCACCAGGTCAGTCTCCAGAA
CCGCCAGTGCAAACAAAAAAAAACATCAGTCATTATCAGAAAATATGCGGAAGGTGACTACTCTAAACTTCTAGAG
GGAGCAACTTTGCGTTTAACAGGGGAAGATATCCTAGATTTTCAAGAAAAAGTCTTCCAAAGTAATGGAACAGGA
GAAAAGATTGAATTATCAAATGGGACTTATACCTTAACAGAAACATCATCTCCAGATGGATATAAAATTGCGGAG
CCGATTAAGTTTAGAGTAGTGAATAAAAAAGTATTTATCGTCCAAAAAGATGGTTCTCAAGTGGAAAATCCAAAC
AAAGAAGTAGCAGAGCCATACTCAGTGGAAGCGTACAGCGATATGCAAGATAGTAACTATATTAATCCAGAAACG
TTCACTCCTTATGGGAAATTTTATTACGCTAAAAATAAGGATAAAAGTTCACAAGTTGTCTACTGTTTTAATGCT
GATTTACACTCTCCACCTGAATCAGAGGATGGGGGAGGAACTATAGATCCTGATATTAGTACGATGAAAGAAGTC
AAGTACACACATACGGCAGGTAGTGATTTGTTTAAATACGCGCTAAGACCGAGAGATACAAATCCAGAAGACTTC
TTAAAGCACATTAAAAAAGTAATTGAAAAAGGCTACAATAAAAAAGGTGATAGCTATAATGGATTAACAGAAACA
CAGTTTCGCGCGGCTACTCAGCTTGCTATCTATTACTTTACAGACAGCACTGACTTAAAAACCTTAAAAACTTAT
AACAATGGGAAAGGTTACCATGGATTTGAATCTATGGATGAAAAAACCCTAGCTGTAACAAAAGAATTAATTAAT
TACGCTCAAGATAATAGTGCCCCTCAACTAACAAATCTTGATTTCTTCGTACCTAATAATAGCAAATACCAATCT
CTTATTGGGACAGAATACCATCCAGATGATTTGGTTGACGTGATTCGTATGGAAGATAAAAAGCAAGAAGTTATT
CCAGTAACTCACAGTTTGACAGTGAAAAAAACAGTAGTCGGTGAGTTGGGAGATAAAACTAAAGGCTTCCAATTT
GAACTTGAGTTGAAAGATAAAACTGGACAGCCTATTGTTAACACTCTAAAAACTAATAATCAAGATTTAGTAGCT
AAAGATGGGAAATATTCATTTAATCTAAAGCATGGTGACACCATAAGAATAGAAGGATTACCGACGGGATATTCT
TATACTCTGAAAGAGACTGAAGCTAAGGATTATATAGTAACCGTTGATAACAAAGTTAGTCAAGAAGCTCAATCA
GCAAGTGAGAATGTCACAGCAGACAAAGAAGTCACTTTTGAAAACCGTAAAGATCTTGTCCCACCAACTGGTTTT
ATTACTGATGGTGGAACCTATCTGTGGTTATTATTGCTTGTCCCATTTGGTTTGTTAGTGTGGTTCTTTGGTCGT
AAAGGACTAAAAAATGACTAA

SEQ ID NO: 109

MNNKKLQKKQDAPRVSNRKPKQLTVTLVGVFLMFLTLVSSMRGAQSIFGEEKRIEEVSVPKIKSPDDAYPWYGYD
SYDSSHPYYERFKVAHDLRVNLNGSKSYQVYCFNINSHYPNRKNAFSKQWFKRVDGTGDVFTNYAQTPKIRGESL
NNKLLSIMYNAYPKNANGYMDKIEPLNAILVTQQAVWYYSDSSYGNIKTLWASELKDGKIDFEQVKLMREAYSKL
ISDDLEETSKNKLPQGSKLNIFVPQDKSVQNLLSAEYVPESPPAPGQSPEPPVQTKKTSVIIRKYAEGDYSKLLE
GATLRLTGEDILDFQEKVFQSNGTGEKIELSNGTYTLTETSSPDGYKIAEPIKFRVVNKKVFIVQKDGSQVENPN
KEVAEPYSVEAYSDMQDSNYINPETFTPYGKFYYAKNKDKSSQVVYCFNADLHSPPESEDGGGTIDPDISTMKEV
KYTHTAGSDLFKYALRPRDTNPEDFLKHIKKVIEKGYNKKGDSYNGLTETQFRAATQLAIYYFTDSTDLKTLKTY
NNGKGYHGFESMDEKTLAVTKELINYAQDNSAPQLTNLDFFVPNNSKYQSLIGTEYHPDDLVDVIRMEDKKQEVI
PVTHSLTVKKTVVGELGDKTKGFQFELELKDKTGQPIVNTLKTNNQDLVAKDGKYSFNLKHGDTIRIEGLPTGYS
YTLKETEAKDYIVTVDNKVSQEAQSASENVTADKEVTFENRKDL*VPPTG*FITDGGTYLWLLLLVPFGLLVWFFGR
KGLKND

19224135 contains an amino acid motif indicative of a cell wall anchor: **SEQ ID NO: 184** VPPTG (shown in italics in SEQ ID NO: 109, above). In some recombinant host cell systems, it may be preferable to remove this motif to facilitate secretion of a recombinant 19224135 protein from the host cell. Alternatively, in other recombinant host cell systems, it may be preferable to use the cell wall anchor motif to anchor the recombinantly expressed protein to the cell wall. The extracellular domain of the expressed protein may be cleaved during purification or the recombinant protein may be left attached to either inactivated host cells or cell membranes in the final composition.

A pilin motif, discussed above, containing a conserved lysine (K) residue has also been identified in 19224135. The pilin motif sequence is underlined in SEQ ID NO: 109, below. Conserved lysine (K) residues are also marked in bold, at amino acid residues 164 and 172. The pilin sequence, in particular the conserved lysine residues, are thought to be important for the formation of oligomeric, pilus-like structures. Preferred fragments of 19224135 include at least one conserved lysine residue. Preferably, fragments include the pilin sequence.

SEQ ID NO: 109

MNNKKLQKKQDAPRVSNRKPKQLTVTLVGVFLMFLTLVSSMRGAQSIFGEEKRIEEVSVPKIKSPDDAYPWYGYD
SYDSSHPYYERFKVAHDLRVNLNGSKSYQVYCFNINSHYPNRKNAFSKQWFKRVDGTGDVFTNYAQTPKIRGESL
NNKLLSIMYNAYP**K**NANGYMD**K**IEPLNAILVTQQAVWYYSDSSYGNIKTLWASELKDGKIDFEQVKLMREAYSKL
ISDDLEETSKNKLPQGSKLNIFVPQDKSVQNLLSAEYVPESPPAPGQSPEPPVQTKKTSVIIRKYAEGDYSKLLE
GATLRLTGEDILDFQEKVFQSNGTGEKIELSNGTYTLTETSSPDGYKIAEPIKFRVVNKKVFIVQKDGSQVENPN
KEVAEPYSVEAYSDMQDSNYINPETFTPYGKFYYAKNKDKSSQVVYCFNADLHSPPESEDGGGTIDPDISTMKEV
KYTHTAGSDLFKYALRPRDTNPEDFLKHIKKVIEKGYNKKGDSYNGLTETQFRAATQLAIYYFTDSTDLKTLKTY
NNGKGYHGFESMDEKTLAVTKELINYAQDNSAPQLTNLDFFVPNNSKYQSLIGTEYHPDDLVDVIRMEDKKQEVI
PVTHSLTVKKTVVGELGDKTKGFQFELELKDKTGQPIVNTLKTNNQDLVAKDGKYSFNLKHGDTIRIEGLPTGYS
YTLKETEAKDYIVTVDNKVSQEAQSASENVTADKEVTFENRKDLVPPTGFITDGGTYLWLLLLVPFGLLVWFFGR
KGLKND

An E box containing a conserved glutamic residue has been identified in 19224135. The E-box motif is underlined in SEQ ID NO: 109, below. The conserved glutamic acid (E), at amino acid residue 339, is marked in bold. The E box motif, in particular the conserved glutamic acid residue, is thought to be important for the formation of oligomeric pilus-like structures of 19224135. Preferred fragments of 19224135 include the conserved glutamic acid residue. Preferably, fragments include the E box motif.

SEQ ID NO: 109

MNNKKLQKKQDAPRVSNRKPKQLTVTLVGVFLMFLTLVSSMRGAQSIFGEEKRIEEVSVPKIKSPDDAYPWYGYD
SYDSSHPYYERFKVAHDLRVNLNGSKSYQVYCFNINSHYPNRKNAFSKQWFKRVDGTGDVFTNYAQTPKIRGESL
NNKLLSIMYNAYPKNANGYMDKIEPLNAILVTQQAVWYYSDSSYGNIKTLWASELKDGKIDFEQVKLMREAYSKL
ISDDLEETSKNKLPQGSKLNIFVPQDKSVQNLLSAEYVPESPPAPGQSPEPPVQTKKTSVIIRKYAEGDYSKLLE
GATLRLTGEDILDFQEKVFQSNGTGEKIELSNGTYTLT**E**TSSPDGYKIAEPIKFRVVNKKVFIVQKDGSQVENPN
KEVAEPYSVEAYSDMQDSNYINPETFTPYGKFYYAKNKDKSSQVVYCFNADLHSPPESEDGGGTIDPDISTMKEV
KYTHTAGSDLFKYALRPRDTNPEDFLKHIKKVIEKGYNKKGDSYNGLTETQFRAATQLAIYYFTDSTDLKTLKTY
NNGKGYHGFESMDEKTLAVTKELINYAQDNSAPQLTNLDFFVPNNSKYQSLIGTEYHPDDLVDVIRMEDKKQEVI
PVTHSLTVKKTVVGELGDKTKGFQFELELKDKTGQPIVNTLKTNNQDLVAKDGKYSFNLKHGDTIRIEGLPTGYS
YTLKETEAKDYIVTVDNKVSQEAQSASENVTADKEVTFENRKDLVPPTGFITDGGTYLWLLLLVPFGLLVWFFGR
KGLKND

19224136 is thought to be a LepA protein. An example of a nucleotide sequence encoding the LepA protein (SEQ ID NO: 110) and a LepA protein amino acid sequence (SEQ ID NO: 111) are set forth below.

SEQ ID NO: 110

ATGACTAATTACCTAAATCGCTTAAATGAGAATCCACTATTTAAAGCTTTCATACGGTTAGTACTTAAGATTTCT
ATTATTGGATTTCTAGGTTACATTCTATTTCAGTATGTTTTTGGCGTCATGATTGTTAACACAAATCAGATGAGT
CCTGCTGTAAGTGCTGGTGATGGAGTCTTATATTATCGTTTGACTGATCGCTATCATATTAATGATGTGGTGGTC
TATGAGGTTGATAACACTTTGAAAGTTGGTCGAATTGCCGCTCAAGCTGGCGATGAGGTTAGTTTTACGCAAGAA
GGAGGACTGTTGATTAATGGGCATCCACCAGAAAAAGAGGTCCCTTACCTGACGTATCCTCACTCAAGTGGTCCA
AACTTTCCCTATAAAGTTCCTACGGGTACGTATTTCATATTGAATGATTATCGTGAAGAACGTTTGGACAGTCGT
TATTATGGGGCGTTACCCATCAATCAAATCAAAGGGAAAATCTCAACTCTATTAAGAGTGAGAGGAATTTAA

SEQ ID NO: 111

MTNYLNRLNENPLFKAFIRLVLKISIIGFLGYILFQYVFGVMIVNTNQMSPAVSAGDGVLYYRLTDRYHINDVVV
YEVDNTLKVGRIAAQAGDEVSFTQEGGLLINGHPPEKEVPYLTYPHSSGPNFPYKVPTGTYFILNDYREERLDSR
YYGALPINQIKGKISTLLRVRGI

19224137 is thought to be a fimbrial protein. An example of a nucleotide sequence encoding the fimbrial protein (SEQ ID NO: 112) and a fimbrial protein amino acid sequence (SEQ ID NO: 113) are set forth below.

SEQ ID NO: 112

ATGAAAAAAAATAAATTATTACTTGCTACTGCAATCTTAGCAACTGCTTTAGGAACAGCTTCTTTAAATCAAAAC
GTAAAAGCTGAGACGGCAGGGGTTGTTAGCAGTGGTCAATTAACAATAAAAAAATCAATTACAAATTTTAATGAT
GATACACTTTTGATGCCTAAGACAGACTATACTTTTAGCGTTAATCCGGATAGTGCGGCTACAGGTACTGAAAGT
AATTTACCAATTAAACCAGGTATTGCTGTTAACAATCAAGATATTAAGGTTTCTTATTCTAATACTGATAAGACA
TCAGGTAAAGAAAAACAAGTTGTTGTTGACTTTATGAAAGTTACTTTTCCTAGCGTTGGTATTTACCGTTATGTT
GTTACCGAGAATAAAGGGACAGCAGAAGGAGTTACATATGATGATACAAAATGGTTAGTTGACGTCTATGTTGGT
AATAATGAAAAGGGAGGTCTTGAACCAAAGTATATTGTATCTAAAAAAGGAGATTCTGCTACTAAAGAACCAATC
CAGTTTAATAATTCATTCGAAACAACGTCATTAAAAATTGAAAAGGAAGTTACTGGTAATACAGGAGATCATAAA
AAAGCATTTACCTTTACATTAACATTGCAACCAAATGAATACTATGAGGCAAGTTCGGTTGTGAAAATTGAAGAG
AACGGACAAACGAAAGATGTGAAAATTGGGGAGGCATATAAGTTTACTTTGAACGATAGTCAGAGTGTGATATTG
TCTAAATTACCAGTTGGTATTAATTATAAAGTTGAAGAAGCAGAAGCTAATCAAGGTGGATATACTACAACAGCA
ACTTTAAAAGATGGAGAAAAGTTATCTACTTATAACTTAGGTCAGGAACATAAAACAGACAAGACTGCTGATGAA
ATCGTTGTCACAAATAACCGTGACACTCAAGTTCCAACTGGTGTTGTAGGCACCCTTGCTCCATTTGCAGTTCTT
AGCATTGTGGCTATTGGTGGAGTTATCTATATTACAAAACGTAAAAAAGCTTAA

SEQ ID NO: 113

MKKNKLLLATAILATALGTASLNQNVKAETAGVVSSGQLTIKKSITNFNDDTLLMPKTDYTFSVNPDSAATGTES
NLPIKPGIAVNNQDIKVSYSNTDKTSGKEKQVVVDFMKVTFPSVGIYRYVVTENKGTAEGVTYDDTKWLVDVYVG
NNEKGGLEPKYIVSKKGDSATKEPIQFNNSFETTSLKIEKEVTGNTGDHKKAFTFTLTLQPNEYYEASSVVKIEE
NGQTKDVKIGEAYKFTLNDSQSVILSKLPVGINYKVEEAEANQGGYTTTATLKDGEKLSTYNLGQEHKTDKTADE
IVVTNNRDT*QVPTG*VVGTLAPFAVLSIVAIGGVIYITKRKKA

19224137 contains an amino acid motif indicative of a cell wall anchor: **SEQ ID NO: 140** QVPTG (shown in italics in SEQ ID NO: 113, above). In some recombinant host cell systems, it may be preferable to remove this motif to facilitate secretion of a recombinant 19224137 protein from the host cell. Alternatively, in other recombinant host cell systems, it may be preferable to use the cell wall anchor motif to anchor the recombinantly expressed protein to the cell wall. The extracellular domain of the expressed protein may be cleaved during purification or the recombinant protein may be left attached to either inactivated host cells or cell membranes in the final composition.

A pilin motif, discussed above, containing a conserved lysine (K) residue has also been identified in 19224137. The pilin motif sequence is underlined in SEQ ID NO: 113, below. A conserved lysine (K) residue is also marked in bold, at amino acid residue 160. The pilin sequence, in particular the conserved lysine residues, are thought to be important for the formation of oligomeric, pilus-like structures. Preferred fragments of 19224137 include the conserved lysine residue. Preferably, fragments include the pilin sequence.

SEQ ID NO: 113

MKKNKLLLATAILATALGTASLNQNVKAETAGVVSSGQLTIKKSITNFNDDTLLMPKTDYTFSVNPDSAATGTES
NLPIKPGIAVNNQDIKVSYSNTDKTSGKEKQVVVDFMKVTFPSVGIYRYVVTENKGTAEGVTYDDTKWLVDVYVG
NNEKGGLEP**K**YIVSKKGDSATKEPIQFNNSFETTSLKIEKEVTGNTGDHKKAFTFTLTLQPNEYYEASSVVKIEE
NGQTKDVKIGEAYKFTLNDSQSVILSKLPVGINYKVEEAEANQGGYTTTATLKDGEKLSTYNLGQEHKTDKTADE
IVVTNNRDTQVPTGVVGTLAPFAVLSIVAIGGVIYITKRKKA

An E box containing a conserved glutamic residue has been identified in 19224137. The E-box motif is underlined in SEQ ID NO: 113, below. The conserved glutamic acid (E), at amino acid residue 263, is marked in bold. The E box motif, in particular the conserved glutamic acid residue, is thought to be important for the formation of oligomeric pilus-like structures of 19224137. Preferred fragments of 19224137 include the conserved glutamic acid residue. Preferably, fragments include the E box motif.

SEQ ID NO: 113

MKKNKLLLATAILATALGTASLNQNVKAETAGVVSSGQLTIKKSITNFNDDTLLMPKTDYTFSVNPDSAATGTES
NLPIKPGIAVNNQDIKVSYSNTDKTSGKEKQVVVDFMKVTFPSVGIYRYVVTENKGTAEGVTYDDTKWLVDVYVG
NNEKGGLEPKYIVSKKGDSATKEPIQFNNSFETTSLKIEKEVTGNTGDHKKAFTFTLTLQPNEYYEASSVVKIEE
NGQTKDVKIGEAYKFTLNDSQSVILSKLPVGINYKVEEAEANQGGYTTTATLKDGEKLSTYNLGQEHKTDKTADE
IVVTNNRDTQVPTGVVGTLAPFAVLSIVAIGGVIYITKRKKA

19224138 is thought to be a SrtC2-type sortase. An example of a nucleotide sequence encoding the SrtC2 sortase (SEQ ID NO: 114) and a SrtC2 sortase amino acid sequence (SEQ ID NO: 115) are set forth below.

SEQ ID NO: 114

ATGATGATGACAATTGTACAGGTTATCAATAAAGCCATTGATACTCTCATTCTTATCTTTTGTTTAGTCGTACTA
TTTTTAGCTGGTTTTGGTTTGTGGGATTCTTATCATCTCTATCAACAAGCAGACGCTTCTAATTTCAAAAAATTT
AAAACAGCTCAACAACAGCCTAAATTTGAAGACTTGTTAGCTTTGAATGAGGATGTCATTGGTTGGTTAAATATC
CCGGGGACTCATATTGATTATCCTCTAGTTCAGGGAAAAACGAATTTAGAGTATATTAATAAAGCAGTTGATGGC
AGTGTTGCCATGTCTGGTAGTTTATTTTTAGATACACGGAATCATAATGATTTTACGGACGATTACTCTCTGATT
TATGGCCATCATATGGCAGGTAATGCCATGTTTGGCGAAATTCCAAAATTTTTAAAAAAGGATTTTTTCAACAAA
CATAATAAAGCTATCATTGAAACAAAAGAGAGAAAAAAACTAACCGTCACTATTTTTGCTTGTCTCAAGACAGAT
GCCTTTGACCAGTTAGTTTTTAATCCTAATGCTATTACCAATCAAGACCAACAAAGGCAGCTCGTTGATTATATC
AGTAAAAGATCAAAACAATTTAAACCTGTTAAATTGAAGCATCATACAAAGTTCGTTGCTTTTTCAACGTGTGAA
AATTTTTCTACTGACAATCGTGTTATCGTTGTCGGTACTATTCAAGAATAA

SEQ ID NO: 115

MMMTIVQVINKAIDTLILIFCLVVLFLAGFGLWDSYHLYQQADASNFKKFKTAQQQPKFEDLLALNEDVIGWLNI
PGTHIDYPLVQGKTNLEYINKAVDGSVAMSGSLFLDTRNHNDFTDDYSLIYGHHMAGNAMFGEIPKFLKKDFFNK
HNKAIIETKERKKLTVTIFACLKTDAFDQLVFNPNAITNQDQQRQLVDYISKRSKQFKPVKLKHHTKFVAFSTCE
NFSTDNRVIVVGTIQE

19224139 is an open reading frame that encodes a sortase substrate motif LPXAG shown in italics in SEQ ID NO: 117. An example of a nucleotide sequence of the open reading frame (SEQ ID NO: 116) and the amino acid sequence encoded by the open reading frame (SEQ ID NO: 117) are set forth below.

SEQ ID NO: 116

ATGTTATTTTCTGTCGTAATGATATTAACCATGCTGGCCTTTAATCAGACTGTTTTAGCAAAAGACAGCACTGTT
CAAACTAGCATTAGTGTCGAAAATGTCTTAGAGAGAGCAGGCGATAGTACCCCATTTTCGATTGCATTAGAATCA
ATTGATGCGATGAAAACAATAGAAGAAATAACAATTGCTGGTTCTGGAAAAGCAAGCTTTTCCCCTCTGACCTTC
ACAACAGTTGGGCAATATACTTATCGTGTTTATCAGAAGCCTTCACAAAATAAAGATTATCAAGCAGATACTACT
GTATTTGACGTTCTTGTCTATGTGACCTATGATGAAGATGGGACTCTAGTCGCAAAAGTTATTTCTCGAAGGGCT
GGAGACGAAGAAAAATCAGCGATTACTTTTAAGCCCAAACGGTTAGTAAAACCAATACCGCCTAGACAACCTAAC
ATCCCTAAAACCCCATTACCATTAGCTGGTGAAGTAAAAAGTTTATTGGGTATCTTAAGTATCGTATTACTGGGG
TTACTAGTTCTTCTTTATGTTAAAAAACTGAAGAG

SEQ ID NO: 117

MLFSVVMILTMLAFNQTVLAKDSTVQTSISVENVLERAGDSTPFSIALESIDAMKTIEEITIAGSGKASFSPLTF
TTVGQYTYRVYQKPSQNKDYQADTTVFDVLVYVTYDEDGTLVAKVISRRAGDEEKSAITFKPKRLVKPIPPRQPN
IPKTP*LPLAG*EVKSLLGILSIVLLGLLVLLYVKKLKSKL

19224139 contains an amino acid motif indicative of a cell wall anchor: **SEQ ID NO: 185** LPLAG (shown in italics in SEQ ID NO: 117, above). In some recombinant host cell systems, it may be preferable to remove this motif to facilitate secretion of a recombinant 19224139 protein from the host cell. Alternatively, in other recombinant host cell systems, it may be preferable to use the cell wall anchor motif to anchor the recombinantly expressed protein to the cell wall. The extracellular domain of the expressed protein may be cleaved during purification or the recombinant protein may be left attached to either inactivated host cells or cell membranes in the final composition.

A pilin motif, discussed above, containing a conserved lysine (K) residue has also been identified in 19224139. The pilin motif sequence is underlined in SEQ ID NO: 117, below. A conserved lysine (K) residue is also marked in bold, at amino acid residue 138. The pilin sequence, in particular the conserved lysine residue, is thought to be important for the formation of oligomeric, pilus-like structures. Preferred fragments of 19224139 include the conserved lysine residue. Preferably, fragments include the pilin sequence.

SEQ ID NO: 117

MLFSVVMILTMLAFNQTVLAKDSTVQTSISVENVLERAGDSTPFSIALESIDAMKTIEEITIAGSGKASFSPLTF
TTVGQYTYRVYQKPSQNKDYQADTTVFDVLVYVTYDEDGTLVAKVISRRAGDEEKSAITFKP**K**RLVKPIPPRQPN
IPKTPLPLAGEVKSLLGILSIVLLGLLVLLYVKKLKSKL

Two E boxes containing conserved glutamic residues have been identified in 19224139. The E-box motifs are underlined in SEQ ID NO: 117, below. The conserved glutamic acid (E) residues, at amino acid residues 58 and 128, are marked in bold. The E box motifs, in particular the conserved glutamic acid residues, are thought to be important for the formation of oligomeric pilus-like structures of 19224139. Preferred fragments of 19224139 include at least one conserved glutamic acid residue. Preferably, fragments include at least one E box motif.

SEQ ID NO: 117

MLFSVVMILTMLAFNQTVLAKDSTVQTSISVENVLERAGDSTPFSIALESIDAMKTI**E**EITIAGSGKASFSPLTF
TTVGQYTYRVYQKPSQNKDYQADTTVFDVLVYVTYDEDGTLVAKVISRRAGD**E**EKSAITFKPKRLVKPIPPRQPN
IPKTPLPLAGEVKSLLGILSIVLLGLLVLLYVKKLKSKL

19224140 is thought to be a MsmRL protein. An example of a nucleotide sequence encoding the MsmRL protein (SEQ ID NO: 118) and a MsmRL protein amino acid sequence (SEQ ID NO: 119) are set forth below.

SEQ ID NO: 118

ATGGTTATATTCGATTTAAAACATGTGCAAACATTACACAGCTTGTCTCAATTACCTATTTCAGTGATGTCACAA
GATAAGGCACTTATTCAAGTATATGGTAATGACGACTATTTATTATGTTACTATCAATTTTTAAAGCATCTAGCT
ATTCCTCAAGCTGCACAAGATGTTATTTTTTATGAGGGTTTATTTGAAGAGTCCTTTATGATTTTTCCTCTTTGT
CACTACATTATTGCCATTGGACCTTTCTACCCTTATTCACTTAATAAAGACTATCAGGAACAATTAGCTAATAAT
TTTTTAAAACATTCTTCTCATCGTAGCAAAGAAGAGCTCTTATCCTATATGGCACTTGTCCCACATTTTCCAATT
AATAATGTGCGGAACCTTTTGATAGCTATTGACGCTTTTTTTGACACACAATTTGAGACGACTTGCCAACAAACA
ATTCATCAATTGTTGCAGCATTCAAAACAGATGACTGCTGATCCTGATATCATTCATCGCCTTAAGCATATTAGC
AAAGCATCTAGCCAACTACCGCCTGTTTTAGAGCACCTAAATCATATTATGGATCTGGTAAAGCTAGGCAATCCA
CAATTGCTCAAGCAAGAAATCAATCGCATCCCCTTATCAAGTATCACCTCATCTTCTATTTCTGCTCTAAGGGCG
GAAAAGAACCTCACTGTTATCTATTTAACTAGGTTACTGGAATTCAGTTTTGTAGAAAATACTGACGTAGCAAAG
CATTATAGCCTTGTCAAATACTACATGGCCTTAAATGAAGAAGCGAGTGACTTGCTCAAAGTTTTGAGAATTCGC
TGTGCAGCCATCATCCATTTTTCCGAATCATTAACCAATAAAAGTATTTCTGATAAACGTCAAATGTACAATAGT
GTGCTTCATTATGTCGATAGTCACCTGTATTCCAAATTAAAGGTATCTGATATCGCTAAGCGCCTATATGTTTCC
GAATCTCACTTACGTTCAGTCTTTAAAAAATACTCAAATGTTTCCTTACAACATTATATTCTAAGTACAAAAATC
AAAGAAGCTCAACTACTCTTAAAACGAGGAATTCCTGTTGGAGAAGTGGCTAAAAGCTTATATTTTTATGACACT
ACCCATTTTCATAAAATCTTTAAAAAATACACGGGTATTTCTTCAAAAGACTATCTTGCTAAATACCGAGATAAT
ATTTAA

SEQ ID NO: 119

MVIFDLKHVQTLHSLSQLPISVMSQDKALIQVYGNDDYLLCYYQFLKHLAIPQAAQDVIFYEGLFEESFMIFPLC
HYIIAIGPFYPYSLNKDYQEQLANNFLKHSSHRSKEELLSYMALVPHFPINNVRNLLIAIDAFFDTQFETTCQQT

IHQLLQHSKQMTADPDILHRLKHISKASSQLPPVLEHLNHIMDLVKLGNPQLLKQEINRIPLSSITSSSISALRA
EKNLTVIYLTRLLEFSFVENTDVAKHYSLVKYYMALNEEASDLLKVLRIRCAAIIHFSESLTNKSISDKRQMYNS
VLHYVDSHLYSKLKVSDIAKRLYVSESHLRSVFKKYSNVSLQHYILSTKIKEAQLLLKRGIPVGEVAKSLYFYDT
THFHKIFKKYTGISSKDYLAKYRDNI

19224141 is thought to be a protein F2 fibronectin binding protein. An example of a nucleotide sequence encoding the protein F2 fibronectin binding protein (SEQ ID NO: 120) and a protein F2 fibronectin binding protein amino acid sequence (SEQ ID NO: 121) are set forth below.

SEQ ID NO: 120

ATGACACAAAAAAATAGCTATAAGTTAAGCTTCCTGTTATCCCTAACAGGATTTATTTTAGGTTTATTATTGGTT
TTTATAGGATTGTCCGGAGTATCAGTAGGACATGCGGAAACAAGAAATGGAGCAAACAAACAAGGATCTTTTGAA
ATCAAGAAAGTCGACCAAAACAATAAGCCTTTACCGGGAGCAACTTTTTCACTGACATCAAAGGATGGCAAGGGA
ACATCTGTTCAAACGTTCACTTCAAATGATAAAGGTATTGTAGATGCTCAAAATCTCCAACCAGGGACTTATACC
TTAAAAGAAGAAACAGCACCAGATGGTTATGATAAAACCAGCCGGACTTGGACAGTGACTGTTTATGAGAACGGC
TATACCAAGTTGGTTGAAAATCCCTATAATGGGGAAATCATCAGTAAAGCAGGGTCAAAAGATGTTAGTAGTTCT
TTACAGTTGGAAAATCCCAAAATGTCAGTTGTTTCTAAATATGGGAAAACAGAGGTTAGTAGTGGCGCAGCGGAT
TTCTACCGCAACCATGCCGCCTATTTTAAAATGTCTTTTGAGTTGAAACAAAAGGATAAATCTGAAACAATCAAC
CCAGGTGATACCTTTGTGTTACAGCTGGATAGACGTCTCAATCCTAAAGGTATCAGTCAAGATATCCCTAAAATC
ATTTACGACAGTGCAAATAGTCCGCTTGCGATTGGAAAATACCATGCTGAGAACCATCAACTTATCTATACTTTC
ACAGATTATATTGCGGGTTTAGATAAAGTCCAGTTGTCTGCAGAATTGAGCTTATTCCTAGAGAATAAGGAAGTG
TTGGAAAATACTAGTATCTCAAATTTTAAGAGTACCATAGGTGGGCAGGAGATCACCTATAAAGGAACGGTTAAT
GTTCTTTATGGAAATGAGAGCACTAAAGAAAGCAATTATATTACTAATGGATTGAGCAATGTGGGTGGGAGTATT
GAAAGCTACAACACCGAAACGGGAGAATTTGTCTGGTATGTTTATGTCAATCCAAACCGTACCAATATTCCTTAT
GCGACCATGAATTTATGGGGATTTGGAAGGGCTCGTTCAAATACAAGCGACTTAGAAAACGACGCTAATACAAGT
AGTGCTGAGCTTGGAGAGATTCAGGTCTATGAAGTACCTGAAGGAGAAAAATTACCATCAAGTTATGGGGTTGAT
GTTACAAAACTTACTTTAAGAACGGATATCACAGCAGGCCTAGGAAATGGTTTTCAAATGACCAAACGTCAGCGA
ATTGACTTTGGAAATAATATCCAAAATAAAGCATTTATCATCAAAGTAACAGGGAAAACAGACCAATCTGGTAAG
CCATTGGTTGTTCAATCCAATTTGGCAAGTTTTCGTGGTGCTTCTGAATATGCTGCTTTTACTCCAGTTGGAGGA
AATGTCTACTTCCAAAACGAAATTGCCTTGTCTCCTTCTAAGGGTAGTGGTTCTGGGAAAAGTGAATTTACTAAG
CCCTCTATTACAGTAGCAAATCTAAAACGAGTGGCTCAGCTTCGCTTTAAGAAAATGTCAACTGACAATGTGCCA
TTGCCAGAAGCGGCTTTTGAGCTGCGTTCATCAAATGGTAATAGTCAGAAATTAGAAGCCAGTTCAAACACACAA
GGAGAGGTTCACTTTAAGGACCTGACCTCGGGCACATATGACCTGTATGAAACAAAAGCGCCAAAAGGTTATCAG
CAGGTGACAGAGAAATTGGCGACCGTTACTGTTGATACTACCAAACCTGCTGAGGAAATGGTCACTTGGGGAAGC
CCACATTCGTCTGTAAAAGTAGAAGCTAACAAAGAAGTCACGATTGTCAACCATAAAGAAACCCTTACGTTTTCA
GGGAAGAAAATTTGGGAGAATGACAGACCAGATCAACGCCCAGCAAAGATTCAAGTGCAACTGTTGCAAAATGGT
CAAAAGATGCCTAACCAGATTCAAGAAGTAACGAAGGATAACGATTGGTCTTATCACTTCAAAGACTTGCCTAAG
TACGATGCCAAGAATCAGGAGTATAAGTACTCAGTTGAAGAAGTAAATGTTCCAGACGGCTACAAGGTGTCGTAT
TTAGGAAATGATATATTTAACACCAGAGAAACAGAATTTGTGTTTGAACAGAATAACTTTAACCTTGAATTTGGA
AATGCTGAAATAAAAGGTCAATCTGGGTCAAAAATCATTGATGAAGACACGCTAACGTCTTTCAAAGGTAAGAAA
ATTTGGAAAAATGATACGGCAGAAAATCGTCCCCAAGCCATTCAAGTGCAGCTTTATGCTGATGGAGTGGCTGTG
GAAGGTCAAACCAAATTTATTTCTGGCTCAGGTAATGAGTGGTCATTTGAGTTTAAAAACTTGAAGAAGTATAAT
GGAACAGGTAATGACATCATTTACTCAGTTAAAGAAGTAACTGTTCCAACAGGTTATGATGTGACTTACTCAGCT
AATGATATTATTAATACCAAACGTGAGGTTATTACACAACAAGGACCGAAACTAGAGATTGAAGAAACGCTTCCG
CTAGAATCAGGTGCTTCAGGCGGTACCACTACTGTCGAAGACTCACGCCCAGTTGATACCTTATCAGGTTTATCA
AGTGAGCAAGGTCAGTCCGGTGATATGACAATTGAAGAAGATAGTGCTACCCATATTAAATTCTCAAAACGTGAT
ATTGACGGCAAAGAGTTAGCTGGTGCAACTATGGAGTTGCGTGATTCATCTGGTAAAACTATTAGTACATGGATT
TCAGATGGACAAGTGAAAGATTTCTACCTGATGCCAGGAAAATATACATTTGTCGAAACCGCAGCACCAGACGGT
TATGAGATAGCAACTGCTATTACCTTTACAGTTAATGAGCAAGGTCAGGTTACTGTAAATGGCAAAGCAACTAAA
GGTGACACTCATATTGTCATGGTTGATGCTTACAAGCCAACTAAGGGTTCAGGTCAGGTTATTGATATTGAAGAA
AAGCTTCCAGACGAGCAAGGTCATTCTGGTTCAACTACTGAAATAGAAGACAGTAAATCTTCAGACCTTATCATT
GGCGGTCAAGGTGAAGTTGTTGACACAACAGAAGACACACAAAGTGGTATGACGGGCCATTCTGGCTCAACTACT
GAAATAGAAGATAGCAAGTCTTCAGACGTTATCATTGGTGGTCAGGGGCAGGTTGTCGAGACAACAGAGGATACC
CAAACTGGCATGTACGGGGATTCTGGTTGTAAAACGGAAGTCGAAGATACTAAACTAGTACAATCCTTCCACTTT
GATAACAAGGAACCAGAAAGTAACTCTGAGATTCCTAAAAAAGATAAGCCAAAGAGTAATACTAGTTTACCAGCA
ACTGGTGAGAAGCAACATAATATGTTCTTTTGGATGGTTACTTCTTGCTCACTTATTAGTAGTGTTTTTGTAATA
TCACTAAAATCCAAAAAACGCCTATCATCATGTTAA

SEQ ID NO: 121

MTQKNSYKLSFLLSLTGFILGLLLVFIGLSGVSVGHAETRNGANKQGSFEIKKVDQNNKPLPGATFSLTSKDGKG
TSVQTFTSNDKGIVDAQNLQPGTYTLKEETAPDGYDKTSRTWTVTVYENGYTKLVENPYNGEIISKAGSKDVSSS
LQLENPKMSVVSKYGKTEVSSGAADFYRNHAAYFKMSFELKQKDKSETINPGDTFVLQLDRRLNPKGISQDIPKI
IYDSANSPLAIGKYHAENHQLIYTFTDYIAGLDKVQLSAELSLFLENKEVLENTSISNFKSTIGGQEITYKGTVN
VLYGNESTKESNYITNGLSNVGGSIESYNTETGEFVWYVYVNPNRTNIPYATMNLWGFGRARSNTSDLENDANTS
SAELGEIQVYEVPEGEKLPSSYGVDVTKLTLRTDITAGLGNGFQMTKRQRIDFGNNIQNKAFIIKVTGKTDQSGK
PLVVQSNLASFRGASEYAAFTPVGGNVYFQNEIALSPSKGSGSGKSEFTKPSITVANLKRVAQLRFKKMSTDNVP
LPEAAFELRSSNGNSQKLEASSNTQGEVHFKDLTSGTYDLYETKAPKGYQQVTEKLATVTVDTTKPAEEMVTWGS
PHSSVKVEANKEVTIVNHKETLTFSGKKIWENDRPDQRPAKIQVQLLQNGQKMPNQIQEVTKDNDWSYHFKDLPK
YDAKNQEYKYSVEEVNVPDGYKVSYLGNDIFNTRETEFVFEQNNFNLEFGNAEIKGQSGSKIIDEDTLTSFKGKK
IWKNDTAENRPQAIQVQLYADGVAVEGQTKFISGSGNEWSFEFKNLKKYNGTGNDIIYSVKEVTVPTGYDVTYSA
NDIINTKREVITQQGPKLEIEETLPLESGASGGTTTVEDSRPVDTLSGLSSEQGQSGDMTIEEDSATHIKFSKRD
IDGKELAGATMELRDSSGKTISTWISDGQVKDFYLMPGKYTFVETAAPDGYEIATAITFTVNEQGQVTVNGKATK
GDTHIVMVDAYKPTKGSGQVIDIEEKLPDEQGHSGSTTEIEDSKSSDLIIGGQGEVVDTTEDTQSGMTGHSGSTT
EIEDSKSSDVIIGGQGQVVETTEDTQTGMYGDSGCKTEVEDTKLVQSFHFDNKEPESNSEIPKKDKPKSNTS*LPA*
*TG*EKQHNMFFWMVTSCSLISSVFVISLKSKKRLSSC

19224141 contains an amino acid motif indicative of a cell wall anchor: **SEQ ID NO: 181** LPATG (shown in italics in SEQ ID NO: 121, above). In some recombinant host cell systems, it may be preferable to remove this motif to facilitate secretion of a recombinant 19224141 protein from the host cell. Alternatively, in other recombinant host cell systems, it may be preferable to use the cell wall anchor motif to anchor the recombinantly expressed protein to the cell wall. The extracellular domain of the expressed protein may be cleaved during purification or the recombinant protein may be left attached to either inactivated host cells or cell membranes in the final composition.

Two pilin motifs, discussed above, containing conserved lysine (K) residues have also been identified in 19224141. The pilin motif sequences are underlined in SEQ ID NO: 121, below. Conserved lysine (K) residues are also marked in bold, at amino acid residues 157 and 163 and at amino acid residues 216, 224, and 238. The pilin sequence, in particular the conserved lysine residues, are thought to be important for the formation of oligomeric, pilus-like structures. Preferred fragments of 19224141 include at least one conserved lysine residue. Preferably, fragments include at least one pilin sequence.

SEQ ID NO: 121

MTQKNSYKLSFLLSLTGFILGLLLVFIGLSGVSVGHAETRNGANKQGSFEIKKVDQNNKPLPGATFSLTSKDGKG
TSVQTFTSNDKGIVDAQNLQPGTYTLKEETAPDGYDKTSRTWTVTVYENGYTKLVENPYNGEIISKAGSKDVSSS
LQLENP**K**MSVVS**K**YGKTEVSSGAADFYRNHAAYFKMSFELKQKDKSETINPGDTFVLQLDRRLNP**K**GISQDIP**K**I
IYDSANSPLAIG**K**YHAENHQLIYTFTDYIAGLDKVQLSAELSLFLENKEVLENTSISNFKSTIGGQEITYKGTVN
VLYGNESTKESNYITNGLSNVGGSIESYNTETGEFVWYVYVNPNRTNIPYATMNLWGFGRARSNTSDLENDANTS
SAELGEIQVYEVPEGEKLPSSYGVDVTKLTLRTDITAGLGNGFQMTKRQRIDFGNNIQNKAFIIKVTGKTDQSGK
PLVVQSNLASFRGASEYAAFTPVGGNVYFQNEIALSPSKGSGSGKSEFTKPSITVANLKRVAQLRFKKMSTDNVP
LPEAAFELRSSNGNSQKLEASSNTQGEVHFKDLTSGTYDLYETKAPKGYQQVTEKLATVTVDTTKPAEEMVTWGS
PHSSVKVEANKEVTIVNHKETLTFSGKKIWENDRPDQRPAKIQVQLLQNGQKMPNQIQEVTKDNDWSYHFKDLPK
YDAKNQEYKYSVEEVNVPDGYKVSYLGNDIFNTRETEFVFEQNNFNLEFGNAEIKGQSGSKIIDEDTLTSFKGKK
IWKNDTAENRPQAIQVQLYADGVAVEGQTKFISGSGNEWSFEFKNLKKYNGTGNDIIYSVKEVTVPTGYDVTYSA
NDIINTKREVITQQGPKLEIEETLPLESGASGGTTTVEDSRPVDTLSGLSSEQGQSGDMTIEEDSATHIKFSKRD
IDGKELAGATMELRDSSGKTISTWISDGQVKDFYLMPGKYTFVETAAPDGYEIATAITFTVNEQGQVTVNGKATK
GDTHIVMVDAYKPTKGSGQVIDIEEKLPDEQGHSGSTTEIEDSKSSDLIIGGQGEVVDTTEDTQSGMTGHSGSTT
EIEDSKSSDVIIGGQGQVVETTEDTQTGMYGDSGCKTEVEDTKLVQSFHFDNKEPESNSEIPKKDKPKSNTSLPA
TGEKQHNMFFWMVTSCSLISSVFVISLKSKKRLSSC

Two E boxes containing conserved glutamic residues have been identified in 19224141. The E-box motifs are underlined in SEQ ID NO: 121, below. The conserved glutamic acid (E) residues, at amino acid residues 567 and 944, are marked in bold. The E box motifs, in particular the conserved glutamic acid residues, are thought to be important for the formation of oligomeric pilus-like structures of 19224141. Preferred fragments of 19224141 include at least one conserved glutamic acid residue. Preferably, fragments include at least one E box motif.

SEQ ID NO: 121

MTQKNSYKLSFLLSLTGFILGLLLVFIGLSGVSVGHAETRNGANKQGSFEIKKVDQNNKPLPGATFSLTSKDGKG
TSVQTFTSNDKGIVDAQNLQPGTYTLKEETAPDGYDKTSRTWTVTVYENGYTKLVENPYNGEIISKAGSKDVSSS
LQLENPKMSVVSKYGKTEVSSGAADFYRNHAAYFKMSFELKQKDKSETINPGDTFVLQLDRRLNPKGISQDIPKI
IYDSANSPLAIGKYHAENHQLIYTFTDYIAGLDKVQLSAELSLFLENKEVLENTSISNFKSTIGGQEITYKGTVN
VLYGNESTKESNYITNGLSNVGGSIESYNTETGEFVWYVYVNPNRTNIPYATMNLWGFGRARSNTSDLENDANTS
SAELGEIQVYEVPEGEKLPSSYGVDVTKLTLRTDITAGLGNGFQMTKRQRIDFGNNIQNKAFIIKVTGKTDQSGK
PLVVQSNLASFRGASEYAAFTPVGGNVYFQNEIALSPSKGSGSGKSEFTKPSITVANLKRVAQLRFKKMSTDNVP
LPEAAFELRSSNGNSQKLEASSNTQGEVHFKDLTSGTYDLY**E**TKAPKGYQQVTEKLATVTVDTTKPAEEMVTWGS
PHSSVKVEANKEVTIVNHKETLTFSGKKIWENDRPDQRPAKIQVQLLQNGQKMPNQIQEVTKDNDWSYHFKDLPK
YDAKNQEYKYSVEEVNVPDGYKVSYLGNDIFNTRETEFVFEQNNFNLEFGNAEIKGQSGSKIIDEDTLTSFKGKK
IWKNDTAENRPQAIQVQLYADGVAVEGQTKFISGSGNEWSFEFKNLKKYNGTGNDIIYSVKEVTVPTGYDVTYSA
NDIINTKREVITQQGPKLEIEETLPLESGASGGTTTVEDSRPVDTLSGLSSEQGQSGDMTIEEDSATHIKFSKRD
IDGKELAGATMELRDSSGKTISTWISDGQVKDFYLMPGKYTFV**E**TAAPDGYEIATAITFTVNEQGQVTVNGKATK
GDTHIVMVDAYKPTKGSGQVIDIEEKLPDEQGHSGSTTEIEDSKSSDLIIGGQGEVVDTTEDTQSGMTGHSGSTT
EIEDSKSSDVIIGGQGQVVETTEDTQTGMYGDSGCKTEVEDTKLVQSFHFDNKEPESNSEIPKKDKPKSN
TSLPATGEKQHNMFFWMVTSCSLISSVFVISLKSKKRLSSC

As discussed above, applicants have also determined the nucleotide and encoded amino acid sequence of fimbrial structural subunits in several other GAS AI-4 strains of bacteria. Examples of sequences of these fimbrial structural subunits are set forth below.

M12 strain isolate 20010296 is a GAS AI-4 strain of bacteria. 20010296_fimbrial is thought to be a fimbrial structural subunit of M12 strain isolate 20010296. An example of a nucleotide sequence encoding the 20010296_fimbrial protein (SEQ ID NO: 257) and a 20010296_fimbrial protein amino acid sequence (SEQ ID NO: 258) are set forth below.

SEQ ID NO: 257

agcagtggtcaattaacaataaaaaatcaattacaaattttaatgatgatacacttttg
atgcctaagacagactatacttttagcgttaatccggatagtgcggctacaggtactgaa
agtaatttaccaattaaaccaggtattgctgttaacaatcaagatattaaggtttcttat
tctaatactgataagacatcaggtaaagaaaaacaagttgttgttgactttatgaaagtt
acttttcctagcgttggtatttaccgttatgttgttaccgagaataaagggacagcagaa
ggagttacatatgatgatacaaaatggttagttgacgtctatgttggtaataatgaaaag
ggaggtcttgaaccaaagtatattgtatctaaaaaaggagattctgctactaaagaacca
atccagtttaataattcattcgaaacaacgtcattaaaaattgaaaaggaagttactggt
aatacaggagatcataaaaaagcatttaactttacattaacattgcaaccaaatgaatac
tatgaggcaagttcggttgtgaaaattgaagagaacggacaaacgaaagatgtgaaaatt
ggggaggcatataagtttactttgaacgatagtcagagtgtgatattgtctaaattacca
gttggtattaattataaagttgaagaagcagaagctaatcaaggtggatatactacaaca
gcaactttaaaagatggagaaaagttatctacttataacttaggtcaggaacataaaaca
gacaagactgctgatgaaatcgt

SEQ ID NO: 258

SSGQLTIKKSITNFNDDTLLMPKTDYTFSVNPDSAATGTESNLP
IKPGIAVNNQDIKVSYSNTDKTSGKEKQVVVDFMKVTFPSVGIYRYVVTENKGTAEGV
TYDDTKWLVDVYVGNNEKGGLEPKYIVSKKGDSATKEPIQFNNSFETTSLKIEKEVTG
NTGDHKKAFNFTLTLQPNEYYEASSVVKIEENGQTKDVKIGEAYKFTLNDSQSVILSK
LPVGINYKVEEAEANQGGYTTTATLKDGEKLSTYNLGQEHKTDKTADEIV

M12 strain isolate 20020069 is a GAS AI-4 strain of bacteria. 20020069_fimbrial is thought to be a fimbrial structural subunit of M12 strain isolate 20020069. An example of a nucleotide sequence encoding the 20020069_fimbrial protein (SEQ ID NO: 259) and a 20020069_fimbrial protein amino acid sequence (SEQ ID NO: 260) are set forth below.

SEQ ID NO: 259

agcagtggtcaattaacaataaaaaaatcaattacaaattttaatgatgatacacttttg
atgcctaagacagactatacttttagcgttaatccggatagtgcggctacaggtactgaa
agtaatttaccaattaaaccaggtattgctgttaacaatcaagatattaaggtttcttat
tctaatactgataagacatcaggtaaagaaaaacaagttgttgttgactttatgaaagtt
actttttcctagcgttggtatttaccgttatgttgttaccgagaataaagggacagcagaa
ggagttacatatgatgatacaaaatggttagttgacgtctatgttggtaataatgaaaag
ggaggtcttgaaccaaagtatattgtatctaaaaaaggagattctgctactaaagaacca
atccagtttaataattcattcgaaacaacgtcattaaaaattgaaaaggaagttactggt
aatacaggagatcataaaaaagcatttaactttacattaacattgcaaccaaatgaatac
tatgaggcaagttcggttgtgaaaattgaagagaacggacaaacgaaagatgtgaaaatt
ggggaggcatataagtttactttgaacgatagtcagagtgtgatattgtctaaattacca
gttggtattaattataaagttgaagaagcagaagctaatcaaggtggatatactacaaca
gcaactttaaaagatggagaaaagttatctacttataacttaggtcaggaacataaaaca
gacaagactgctgatgaaatcgt

SEQ ID NO: 260

SSGQLTIKKSITNFNDDTLLMPKTDYTFSVNPDSAATGTESNLP
IKPGIAVNNQDIKVSYSNTDKTSGKEKQVVVDFMKVTFPSVGIYRYVVTENKGTAEGV
TYDDTKWLVDVYVGNNEKGGLEPKYIVSKKGDSATKEPIQFNNSFETTSLKIEKEVTG
NTGDHKKAFNFTLTLQPNEYYEASSVVKIEENGQTKDVKIGEAYKFTLNDSQSVILSK
LPVGINYKVEEAEANQGGYTTTATLKDGEKLSTYNLGQEHKTDKTADEIV

M12 strain isolate CDC SS 635 is a GAS AI-4 strain of bacteria. CDC SS 635_fimbrial is thought to be a fimbrial structural subunit of M12 strain isolate CDC SS 635. An example of a nucleotide sequence encoding the CDC SS 635_fimbrial protein (SEQ ID NO: 261) and a CDC SS 635_fimbrial protein amino acid sequence (SEQ ID NO: 262) are set forth below.

SEQ ID NO: 261

gagacggcaggggttgttagcagtggtcaattaacaataaaaaaatcaattacaaatttt
aatgatgatacacttttgatgcctaagacagactatacttttagcgttaatccggatagt
gcggctacaggtactgaaagtaatttaccaattaaaccaggtattgctgttaacaatcaa
gatattaaggtttcttattctaatactgataagacatcaggtaaagaaaaacaagttgtt
gttgactttatgaaagttactttttcctagcgttggtatttaccgttatgttgttaccgag
aataaagggacagcagaaggagttacatatgatgatacaaaatggttagttgacgtctat
gttggtaataatgaaaagggaggtcttgaaccaaagtatattgtatctaaaaaaggagat
tctgctactaaagaaccaatccagtttaataattcattcgaaacaacgtcattaaaaatt
gaaaaggaagttactggtaatacaggagatcataaaaaagcatttaactttacattaaca
ttgcaaccaaatgaatactatgaggcaagttcggttgtgaaaattgaagagaacggacaa
acgaaagatgtgaaaattggggaggcatataagtttactttgaacgatagtcagagtgtg
atattgtctaaattaccagttggtattaattataaagttgaagaagcagaagctaatcaa
ggtggatatactacaacagcaactttaaaagatggagaaaagttatctacttataactta
ggtcaggaacataaaacagacaagactgctgatgaaatcgttgtcacaaataaccgtgac
act

SEQ ID NO: 262

ETAGVVSSGQLTIKKSITNFNDDTLLMPKTDYTFSVNPDSAATG
TESNLPIKPGIAVNNQDIKVSYSNTDKTSGKEKQVVVDFMKVTFPSVGIYRYVVTENK
GTAEGVTYDDTKWLVDVYVGNNEKGGLEPKYIVSKKGDSATKEPIQFNNSFETTSLKI
EKEVTGNTGDHKKAFNFTLTLQPNEYYEASSVVKIEENGQTKDVKIGEAYKFTLNDSQ
SVILSKLPVGINYKVEEAEANQGGYTTTATLKDGEKLSTYNLGQEHKTDKTADEIVVT
NNRDT

M5 strain isolate ISS4883 is a GAS AI-4 strain of bacteria. ISS4883_fimbrial is thought to be a fimbrial structural subunit of M5 strain isolate ISS 4883. An example of a nucleotide sequence encoding the ISS4883_fimbrial protein (SEQ ID NO: 265) and an ISS4883_fimbrial protein amino acid sequence (SEQ ID NO: 266) are set forth below.

SEQ ID NO: 265

```
gagacggcaggggttgtaacaggaaaatcactacaagttacaaagacaatgacttatgat
gatgaagaggtgttaatgcccgaaaccgcctttacttttactatagagcctgatatgact
gcaagtggaaaagaaggcgacctagatattaaaaatggaattgtagaaggcttagacaaa
caagtaacagtaaaatataagaatacagataaaacatctcaaaaaactaaaatagcacaa
tttgatttttctaaggttaaatttccagctataggtgtttaccgctatatggtttcagag
aaaaacgataaaaaagacggaattaggtacgatgataaaaagtggactgtagatgtttat
gttgggaataaggccaataacgaagaaggtttcgaagttctatatattgtatcaaaagaa
ggtacttctagtactaaaaaaccaattgaatttacaaactctattaaaactacttcctta
aaaattgaaaaacaaataactggcaatgcaggagatcgtaaaaaatcattcaacttcaca
ttaacattacaaccaagtgaatattataaaaccggatcagttgtgaaaatcgaacaggat
ggaagtaaaaaagatgtgacgataggaacgccttacaaatttactttgggacacggtaag
agtgtcatgttatcgaaattaccaattggtatcaattactatcttagtgaagacgaagcg
aataaagacggttacactacaacggcaacattaaaagaacaaggcaaagaaaagagttcc
gatttcactttgagtactcaaaaccagaaaacagacgaatctgctgacgaaatcgttgtc
acaaataagcgtgacactctcgag
```

SEQ ID NO: 266

```
ETAGVVTGKSLQVTKTMTYDDEEVLMPETAFTFTIEPDMTASGK
EGDLDIKNGIVEGLDKQVTVKYKNTDKTSQKTKIAQFDFSKVKFPAIGVYRYMVSEKN
DKKDGIRYDDKKWTVDVYVGNKANNEEGFEVLYIVSKEGTSSTKKPIEFTNSIKTTSL
KIEKQITGNAGDRKKSFNFTLTLQPSEYYKTGSVVKIEQDGSKKDVTIGTPYKFTLGH
GKSVMLSKLPIGINYYLSEDEANKDGYTTTATLKEQGKEKSSDFTLSTQNQKTDESAD
EIVVTNKRDTLE
```

M50 strain isolate ISS4538 is a GAS AI-4 strain of bacteria. ISS4538_fimbrial is thought to be a fimbrial structural subunit of M50 strain ISS 4538. An example of a nucleotide sequence encoding the ISS4538_fimbrial protein (SEQ ID NO: 255) and an ISS4538_fimbrial protein amino acid sequence (SEQ ID NO: 256) are set forth below.

SEQ ID NO: 255

```
atgaaaaaaaataaattattacttgctactgcaatcttagcaactgctttaggaacagct
tctttaaatcaaaacgtaaaagctgagacggcaggggttgttagcagtggtcaattaaca
ataaaaaaatcaattacaaattttaatgatgatacacttttgatgcctaagacagactat
acttttagcgttaatccggatagtgcggctacaggtactgaaagtaatttaccaattaaa
ccaggtattgctgttaacaatcaagatattaaggtttcttattctaatactgataagaca
tcaggtaaagaaaaacaagttgttgttgactttatgaaagttacttttcctagcgttggt
atttaccgttatgttgttaccgagaataaagggacagcagaaggagttacatatgatgat
acaaaatggttagttgacgtctatgttggtaataatgaaaagggaggtcttgaaccaaag
tatattgtatctaaaaaggagattctgctactaaagaaccaatccagtttaataattca
ttcgaaacaacgtcattaaaaattgaaaagaaagttactggtaatacaggagatcataaa
aaagcatttaactttacattaacattgcaaccaaatgaatactatgaggcaagttcggtt
gtgaaaattgaagagaacggacaaacgaaagatgtgaaaattggggaggcatataagttt
actttgaacgatagtcagagtgtgatattgtctaaattaccagttggtattaattataaa
gttgaagaagcagaagctaatcaaggtggatatactacaacagcaactttaaaagatgga
gaaaagttatctacttataacttaggtcaggaacataaaacagacaagactgctgatgaa
atcgttgtcacaaataancgngacactcnagttccaacnggtgtngtaggcaccccncct
ccattcncagttcttancattgnggctantggtggngtnatntatnttacaaaacgnaaa
aaagnataa
```

SEQ ID NO: 256

```
MKKNKLLLATAILATALGTASLNQNVKAETAGVVSSGQLTIKKS
ITNFNDDTLLMPKTDYTFSVNPDSAATGTESNLPIKPGIAVNNQDIKVSYSNTDKTSG
```

KEKQVVVDFMKVTFPSVGLYRYVVTENKGTAEGVTYDDTKWLVDVYVGNNEKGGLEPK
YIVSKKGDSATKEPIQFNNSFETTSLKIEKKVTGNTGDHKKAFNFTLTLQPNEYYEAS
SVVKIEENGQTKDVKIGEAYKFTLNDSQSVILSKLPVGINYKVEEAEANQGGYTTTAT
LKDGEKLSTYNLGQEHKTDKTADEIVVTNXRDTXVPTGVVGTPPPFXVLXIXAXGGVX
YXTKRKKX

There may be an upper limit to the number of GAS proteins which will be in the compositions of the invention. Preferably, the number of GAS proteins in a composition of the invention is less than 20, less than 19, less than 18, less than 17, less than 16, less than 15, less than 14, less than 13, less than 12, less than 11, less than 10, less than 9, less than 8, less than 7, less than 6, less than 5, less than 4, or less than 3. Still more preferably, the number of GAS proteins in a composition of the invention is less than 6, less than 5, or less than 4. Still more preferably, the number of GAS proteins in a composition of the invention is 3.

The GAS proteins and polynucleotides used in the invention are preferably isolated, *i.e.*, separate and discrete, from the whole organism with which the molecule is found in nature or, when the polynucleotide or polypeptide is not found in nature, is sufficiently free of other biological macromolecules so that the polynucleotide or polypeptide can be used for its intended purpose.

<u>Examples Other Gram positive bacterial Adhesin Island Sequences</u>

The Gram positive bacteria AI polypeptides of the invention can, of course, be prepared by various means (*e.g.* recombinant expression, purification from a gram positive bacteria, chemical synthesis *etc.*) and in various forms (*e.g.* native, fusions, glycosylated, non-glycosylated *etc.*). They are preferably prepared in substantially pure form (*i.e.* substantially free from other streptococcal or host cell proteins) or substantially isolated form.

The Gram positive bacteria AI proteins of the invention may include polypeptide sequences having sequence identity to the identified Gram positive bacteria proteins. The degree of sequence identity may vary depending on the amino acid sequence (a) in question, but is preferably greater than 50% (*e.g.* 60%, 65%, 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 99.5% or more). Polypeptides having sequence identity include homologs, orthologs, allelic variants and mutants of the identified Gram positive bacteria proteins. Typically, 50% identity or more between two proteins is considered to be an indication of functional equivalence. Identity between proteins is preferably determined by the Smith-Waterman homology search algorithm as implemented in the MPSRCH program (Oxford Molecular), using an affinity gap search with parameters *gap open penalty=12* and *gap extension penalty=1*.

The Gram positive bacteria adhesin island polynucleotide sequences may include polynucleotide sequences having sequence identity to the identified Gram positive bacteria adhesin island polynucleotide sequences. The degree of sequence identity may vary depending on the polynucleotide sequence in question, but is preferably greater than 50% (*e.g.* 60%, 65%, 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 99.5% or more).

The Gram positive bacteria adhesin island polynucleotide sequences of the invention may include polynucleotide fragments of the identified adhesin island sequences. The length of the fragment may vary depending on the polynucleotide sequence of the specific adhesin island sequence, but the fragment is preferably at least 10 consecutive polynucleotides, (*e.g.* at least 10, 12, 14, 16, 18, 20, 25, 30, 35, 40, 50, 60, 70, 80, 90, 100, 150, 200 or more).

The Gram positive bacteria adhesin island amino acid sequences of the invention may include polypeptide fragments of the identified Gram positive bacteria proteins. The length of the fragment may vary depending on the amino acid sequence of the specific Gram positive bacteria antigen, but the fragment is preferably at least 7 consecutive amino acids, (*e.g.* 8, 10, 12, 14, 16, 18, 20, 25, 30, 35, 40, 50, 60, 70, 80, 90, 100, 150, 200 or more). Preferably the fragment comprises one or more epitopes from the sequence. The fragment may comprise at least one T-cell or, preferably, a B-cell epitope of the sequence. T- and B-cell epitopes can be identified empirically (e.g., using PEPSCAN [Geysen *et al.* (1984) *PNAS USA* 81:3998-4002; Carter (1994) *Methods Mol. Biol.* 36:207-223, or similar methods], or they can be predicted (*e.g.*, using the Jameson-Wolf antigenic index [Jameson, BA *et al.* 1988, *CABIOS* 4(1):1818-186], matrix-based approaches [Raddrizzani and Hammer (2000) *Brief Bioinform.* 1(2):179-189], TEPITOPE [De Lalla *et al.* (199) *J. Immunol.* 163:1725-1729], neural networks [Brusic *et al.* (1998) *Bioinformatics* 14(2):121-130], OptiMer & EpiMer [Meister *et al.* (1995) *Vaccine* 13(6):581-591; Roberts *et al.* (1996) *AIDS Res. Hum. Retroviruses* 12(7):593-610], ADEPT [Maksyutov & Zagrebelnaya (1993) *Comput. Appl. Biosci.* 9(3):291-297], Tsites [Feller & de la Cruz (1991) *Nature* 349(6311):720-721], hydrophilicity [Hopp (1993) *Peptide Research* 6:183-190], antigenic index [Welling *et al.* (1985)*FEBS Lett.* 188:215-218] or the methods disclosed in Davenport *et al.* (1995) *Immunogenetics* 42:392-297, etc. Other preferred fragments include (1) the N-terminal signal peptides of each identified Gram positive bacteria protein, (2) the identified Gram positive bacteria protein without their N-terminal signal peptides, (3) each identified Gram positive bacteria protein wherein up to 10 amino acid residues (e.g. 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25 or more) are deleted from the N-terminus and/or the C-terminus *e.g.* the N-terminal amino acid residue may be deleted. Other fragments omit one or more domains of the protein (*e.g.* omission of a signal peptide, of a cytoplasmic domain, of a transmembrane domain, or of an extracellular domain), and (4) the polypeptides, but without their N-terminal amino acid residue.

As indicated in the above text, nucleic acids and polypeptides of the invention may include sequences that:

(a) are identical (*i.e.*, 100% identical) to the sequences disclosed in the sequence listing;
(b) share sequence identity with the sequences disclosed in the sequence listing;
(c) have 1, 2, 3, 4, 5, 6, 7, 8, 9, or 10 single nucleotide or amino acid alterations (deletions, insertions, substitutions), which may be at separate locations or may be contiguous, as compared to the sequences of (a) or (b);
(d) when aligned with a particular sequence from the sequence listing using a pairwise alignment algorithm, a moving window of x monomers (amino acids or nucleotides)